

Sequence length 2175

CACGCGTCCGCAAATTTCTGATTCTTTGAATTAGGATTCCAGATGGGGGCTCATTTCTACAGCCCCAACATTCTCT  
ATAGCCGTTATCACTGCCATCACCAGTCCACCAGCATCTTCTTGAGATTCCACCCCTGCTCCCCAGAGACTTCTCTGC  
TTTGAAGTGAGCAGAAAGGAAGCTCTCAGAAAAATCTCTAGTGGTGGCTGCCGTCGCTCCAGACAATCGGAATCCTGC

M G W L F L K V L L A G V S F S G 17  
CTTCAACCACC ATG GGC TGG CTT TTT CTA AAG GTT TTG TTG GCG GGA GTG AGT TTC TCA GGA 51

F L Y P L V D F C I S G K T R G Q K P N 37  
TTT CTT TAT CCT CTT GTG GAT TTT TGC ATC AGT GGG AAA ACA AGA GGA CAG AAG CCA AAC 111

F V I I L A D D M G W G D L G A N W A E 57  
TTT GTG ATT ATT TTG GCC GAT GAC ATG GGG TGG GGT GAC CTG GGA GCA AAC TGG GCA GAA 171

T K D T A N L D K M A S E G M R F V D F 77  
ACA AAG GAC ACT GCC AAC CTT GAT AAG ATG GCT TCG GAG GGA ATG AGG TTT GTG GAT TTC 231

H A A A S T C S P S R A S L L T G R L G 97  
CAT GCA GCT GCC TCC ACC TGC TCA CCC TCC CGG GCT TCC TTG CTC ACC GGC CGG CTT GGC 291

L R N G V T R N F A V T S V G G L P L N 117  
CTT CGC AAT GGA GTC ACA CGC AAC TTT GCA GTC ACT TCT GTG GGA GGC CTT CCG CTC AAC 351

E T T L A E V L Q Q A G Y V T G I I G K 137  
GAG ACC ACC TTG GCA GAG GTG CTG CAG CAG GCG GGT TAC GTC ACT GGG ATA ATA GGC AAA 411

W H L G H H G S Y H P N F R G F D Y Y F 157  
TGG CAT CTT GGA CAC CAC GGC TCT TAT CAC CCC AAC TTC CGT GGT TTT GAT TAC TAC TTT 471

G I P Y S H D M G C T D T P G Y N H P P 177  
GGA ATC CCA TAT AGC CAT GAT ATG GGC TGT ACT GAT ACT CCA GGC TAC AAC CAC CCT CCT 531

C P A C P Q G D G P S R N L Q R D C Y T 197  
TGT CCA GCG TGT CCA CAG GGT GAT GGA CCA TCA AGG AAC CTT CAA AGA GAC TGT TAC ACT 591

D V A L P L Y E N L N I V E Q P V N L S 217  
GAC GTG GCC CTC CCT CTT TAT GAA AAC CTC AAC ATT GTG GAG CAG CCG GTG AAC TTG AGC 651

S L A Q K Y A E K A T Q F I Q R A S T S 237  
AGC CTT GCC CAG AAG TAT GCT GAG AAA GCA ACC CAG TTC ATC CAG CGT GCA AGC ACC AGC 711

G R P F L L Y V A L A H M H V P L P V T 257  
GGG AGG CCC TTC CTG CTC TAT GTG GCT CTG GCC CAC ATG CAC GTG CCC TTA CCC GTG ACT 771

Q L P A A P R G R S L Y G A G L W E M D 277  
CAG CTA CCA GCA GCG CCA CGG GGC AGA AGC CTG TAT GGT GCA GGG CTC TGG GAG ATG GAC 831

S L V G Q I K D K V D H T V K E N T F L 297  
AGT CTG GTG GGC CAG ATC AAG GAC AAA GTT GAC CAC ACA GTG AAG GAA AAC ACA TTC CTC 891

W F T G D N G P W A Q K C E L A G S V G 317  
TGG TTT ACA GGA GAC AAT GGC CCG TGG GCT CAG AAG TGT GAG CTA GCG GGC AGT GTG GGT 951

P F T G F W Q T R Q G G S P A K Q T T W 337  
CCC TTC ACT GGA TTT TGG CAA ACT CGT CAA GGG GGA AGT CCA GCC AAG CAG ACG ACC TGG 1011

E G G H R V P A L A Y W P G R V P V N V 357

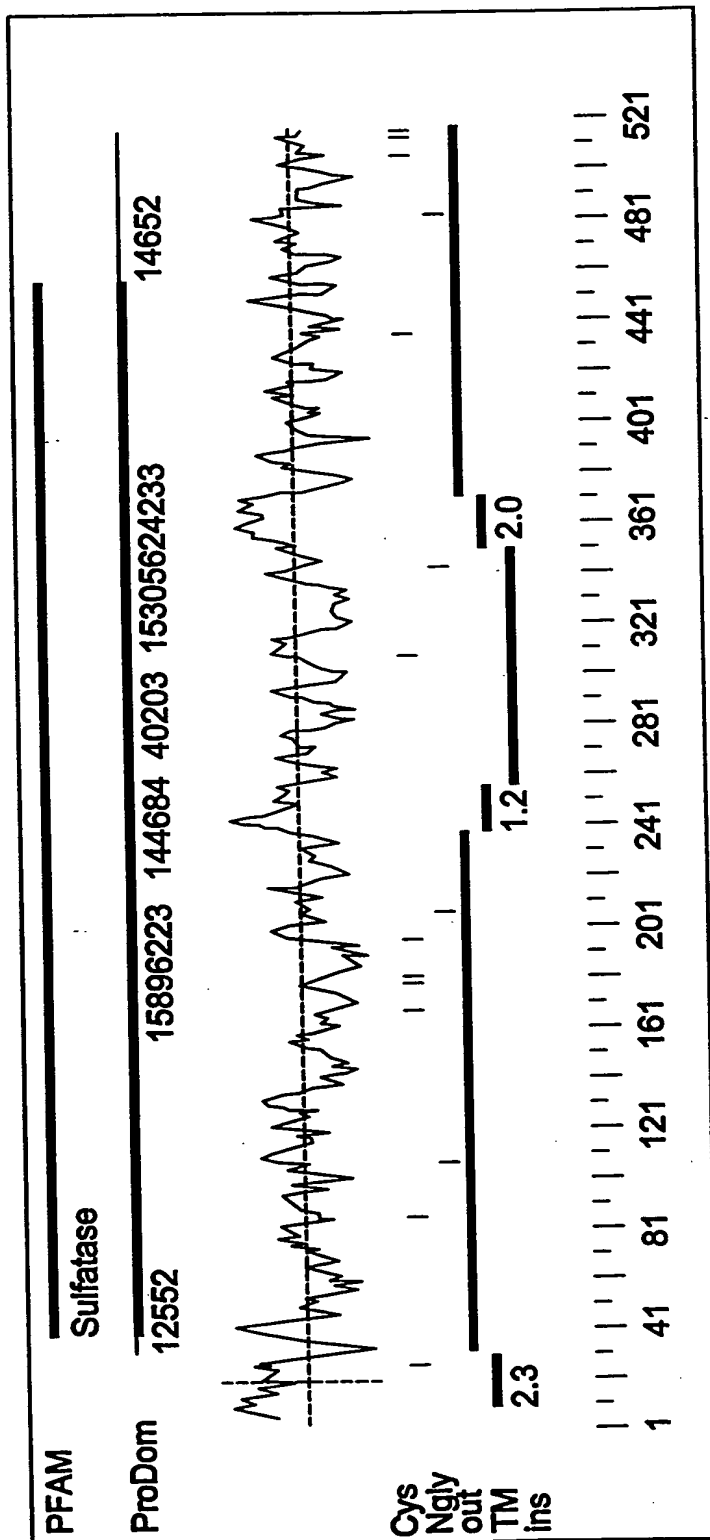
2/44

GAA GGA GGC CAC CGG GTC CCA GCA CTC CCA GCA CTG GCT TAC TGG CCT GGC AGA GTT CCA GTT AAT GTC	1071
T S T A L L S V L D I F P T V V A L A Q	377
ACC AGC ACT GCC TTG TTA AGC GTG CTG GAC ATT TTT CCA ACT GTG GTA GCC CTG GCC CAG	1131
A S L P Q G R R F D G V D V S E V L F G	397
GCC AGC TTA CCT CAA GGA CGG CGC TTT GAT GGT GTG GAC GTC TCC GAG GTG CTC TTT GGC	1191
R S Q P G H R V L F H P N S G A A G E F	417
CGG TCA CAG CCT GGG CAC AGG GTG CTG TTC CAC CCC AAC AGC GGG GCA GCT GGA GAG TTT	1251
G A L Q T V R L E R Y K A F Y I T G G A	437
GGA GCC CTG CAG ACT GTC CGC CTG GAG CGT TAC AAG GCC TTC TAC ATT ACC GGT GGA GCC	1311
R A C D G S T G P E L Q H K F P L I F N	457
AGG GCT TGT GAT GGG AGC ACG GGG CCT GAG CTG CAG CAT AAG TTT CCT CTG ATT TTC AAC	1371
L E D D T A E A V P L E R G G A E Y Q A	477
CTG GAA GAC GAT ACC GCA GAA GCT GTG CCC CTA GAA AGA GGT GGT GCG GAG TAC CAG GCT	1431
V L P E V R K V L A D V L Q D I A N D N	497
GTG CTG CCC GAG GTC AGA AAG GTT CTT GCA GAC GTC CTC CAA GAC ATT GCC AAC GAC AAC	1491
I S S A D Y T Q D P S V T P C C N P Y Q	517
ATC TCC AGC GCA GAT TAC ACT CAG GAC CCT TCA GTA ACT CCC TGC TGT AAT CCC TAC CAA	1551
I A C R C Q A A *	526
ATT GCC TGC CGC TGT CAA GCC GCA TAA	1578

CAGACCAATTTTATTCCACGAGGAGGAGTACCTGGAAATTAGGCAAGTTTGCTTCCAAATTTTATTTTACCCTCTTT  
ACAAACACACGCTTTAGTTTAGTCTTGGAGTTTAGTTTGGAGTTAGCCTTGCAATATCCCTTCTGTATCCTGTCCCTCC  
TCCACGCCGACCCGAGAGCAGCTGAGCTGCGCTGGCTCTGGGCACCCAGTGTGCCTTAATGGGAAGCACACGGGCTTTG  
GAGTCAGGCACAGGTGCCAGCTCCAGCTTTTGAACCTGGGCAATTGTTTAACTAACCTGCAAGTTGATTTTGAGGGTT  
AAATAAAGGCATACATGAAAAAAAAAAAAAAAAA

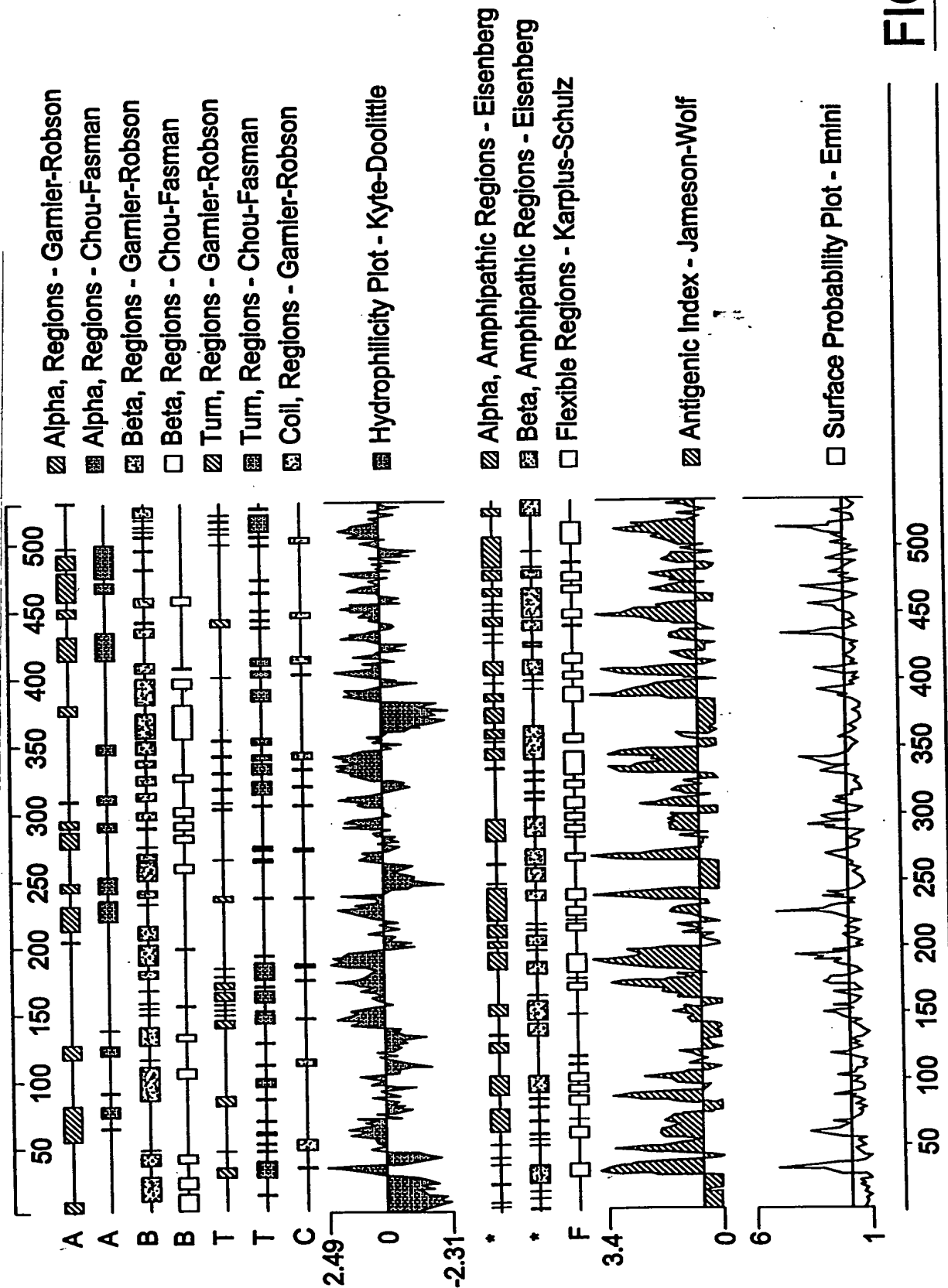
**FIG. 1B.**

3/44



**FIG. 2.**

4/44



5/44

## FIG. 4.

### Prosite Pattern Matches

Prosite version: Release 12.2 of February 1995

>PS00001/PDOC00001/ASN\_GLYCOSYLATION N-glycosylation site.

Query: 117 NETT 120

Query: 215 NLSS 218

Query: 356 NVTs 359

Query: 497 NISS 500

>PS00005/PDOC00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 28 SGK 30

Query: 93 TGR 95

Query: 237 SGR 239

Query: 290 TVK 292

Query: 422 TVR 424

>PS00006/PDOC00006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 120 TLAE 123

Query: 290 TVKE 293

Query: 335 TTWE 338

Query: 364 SVLD 367

Query: 444 TGPE 447

Query: 499 SSAD 502

>PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 12 GVSFSG 17

Query: 33 GQKPNF 38

Query: 52 GANWAE 57

Query: 97 GLRNGV 102

Query: 113 GLPLNE 118

Query: 158 GIPYSH 163

Query: 328 GGSPAK 333

Query: 388 GVDVSE 393

Query: 418 GALQTV 423

Query: 435 GGARAC 440

>PS00009/PDOC00009/AMIDATION Amidation site.

Query: 382 QGRR 385

>PS00149/PDOC00117/SULFATASE\_2 Sulfatases signature 2.

Query: 129 GYVTGIIGKW 138

6/44

**FIG. 5A.**

Input file Fbh23553fl.seq Output File 23553.trans  
Sequence length 4321

```
CCCACGCGTCCGGCTAATGAATCTTGGGGCCGGTGTGGGGCCGGGGCGGCTTGATCGGCAACTAGGAAACCCAGGCGC
AGAGGCCAGGAGCGAGGGCAGGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGCGCTCCTCGGAGGTCAGGGCAGAT
GAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTATCTCTGCTCCTCTGCCTCCTCCCTG
TTCCAGAGCTTTTTCTCTAGAGAAGATTTGAAGGCGGCTTTTGTGCTGACGGCCACCCACCATCATCTAAGAAGAT
AAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAATTGCAGAAAATCTTCAAAGGACCCATCTGCAGATGTT
CTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGGATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGA

                                     M K Y S C C A L V L A      11
CATT TGT CAG TTT TGA CATT GGACCAAATACA ATG AAG TAT TCT TGC TGT GCT CTG GTT TTG GCT      33
V L G T E L L G S L C S T V R S P R F R      31
GTC CTG GGC ACA GAA TTG CTG GGA AGC CTC TGT TCG ACT GTC AGA TCC CCG AGG TTC AGA      93
G R I Q Q E R K N I R P N I I L V L T D      51
GGA CGG ATA CAG CAG GAA CGA AAA AAC ATC CGA CCC AAC ATT ATT CTT GTG CTT ACC GAT      153
D Q D V E L G S L Q V M N K T R K I M E      71
GAT CAA GAT GTG GAG CTG GGG TCC CTG CAA GTC ATG AAC AAA ACG AGA AAG ATT ATG GAA      213
H G G A T F I N A F V T T P M C C P S R      91
CAT GGG GGG GCC ACC TTC ATC AAT GCC TTT GTG ACT ACA CCC ATG TGC TGC CCG TCA CGG      273
S S M L T G K Y V H N H N V Y T N N E N      111
TCC TCC ATG CTC ACC GGG AAG TAT GTG CAC AAT CAC AAT GTC TAC ACC AAC AAC GAG AAC      333
C S S P S W Q A M H E P R T F A V Y L N      131
TGC TCT TCC CCC TCG TGG CAG GCC ATG CAT GAG CCT CGG ACT TTT GCT GTA TAT CTT AAC      393
N T G Y R T A F F G K Y L N E Y N G S Y      151
AAC ACT GGC TAC AGA ACA GCC TTT TTT GGA AAA TAC CTC AAT GAA TAT AAT GGC AGC TAC      453
I P P G W R E W L G L I K N S R F Y N Y      171
ATC CCC CCT GGG TGG CGA GAA TGG CTT GGA TTA ATC AAG AAT TCT CGC TTC TAT AAT TAC      513
I V C R N G I K E K H G F D Y A K D Y F      191
ACT GTT TGT CGC AAT GGC ATC AAA GAA AAG CAT GGA TTT GAT TAT GCA AAG GAC TAC TTC      573
T D L I T N E S I N Y F K M S K R M Y P      211
ACA GAC TTA ATC ACT AAC GAG AGC ATT AAT TAC TTC AAA ATG TCT AAG AGA ATG TAT CCC      633
H R P V M M V I S H A A P H G P E D S A      231
CAT AGG CCC GTT ATG ATG GTG ATC AGC CAC GCT GCG CCC CAC GGC CCC GAG GAC TAC GCC      693
P Q F S K L Y P N A S Q H I T P S Y N Y      251
CCA CAG TTT TCT AAA CTG TAC CCC AAT GCT TCC CAA CAC ATA ACT CCT AGT TAT AAC TAT      753
A P N M D K H W I M Q Y T G P M L P I H      271
GCA CCA AAT ATG GAT AAA CAC TGG ATT ATG CAG TAC ACA GGA CCA ATG CTG CCC ATC CAC      813
M E F T N I L Q R K R L Q T L M S V D D      291
ATG GAA TTT ACA AAC ATT CTA CAG CGC AAA AGG CTC CAG ACT TTG ATG TCA GTG GAT GAT      873
S V E R L Y N M L V E T G E L E N T Y I      311
```

7/44

## FIG. 5B.

```
TCT GTG GAG AGG CTG TAT AAC ATG CTC GTG GAG ACG GGG GAG CTG GAG AAT ACT TAC ATC 933
I Y T A D H G Y H I G Q F G L V K G K S 331
ATT TAC ACC GCC GAC CAT GGT TAC CAT ATT GGG CAG TTT GCA CTG GTC AAG GGG AAA TCC 993
M P Y D F D I R V P F F I R G P S V E P 351
ATG CCA TAT GAC TTT GAT ATT CGT GTG CCT TTT TTT ATT CGT GGT CCA AGT GTA GAA CCA 1053
G S I V P Q I V L N I D L A P T I L D I 371
GGA TCA ATA GTC CCA CAG ATC GTT CTC AAC ATT GAC TTG GCC CCC ACG ATC CTG GAT ATT 1113
A G L D T P P D V D G K S V L K L L D P 391
GCT GGG CTC GAC ACA CCT CCT GAT GTG GAC GGC AAG TCT GTC CTC AAA CTT CTG GAC CCA 1173
E K P G N R F R T N K K A K I W R D T F 411
GAA AAG CCA GGT AAC AGG TTT CGA ACA AAC AAG AAG GCC AAA ATT TGG CGT GAT ACA TTC 1233
L V E R G K F L R K K E E S S K N I Q Q 431
CTA GTG GAA AGA GGC AAA TTT CTA CGT AAG AAG GAA GAA TCC AGC AAG AAT ATC CAA CAG 1293
S N H L P K Y E R V K E L C Q Q A R Y Q 451
TCA AAT CAC TTG CCC AAA TAT GAA CGG GTC AAA GAA CTA TGC CAG CAG GCC AGG TAC CAG 1353
T A C E Q P G Q K W Q C I E D T S G K L 471
ACA GCC TGT GAA CAA CCG GGG CAG AAG TGG CAA TGC ATT GAG GAT ACA TCT GGC AAG CTT 1413
R I H K C K G P S D L L T V R Q S T R N 491
CGA ATT CAC AAG TGT AAA GGA CCC AGT GAC CTG CTC ACA GTC CGG CAG AGC ACG CGG AAC 1473
L Y A R G F H D K D K E C S C R E S G Y 511
CTC TAC GCT CGC GGC TTC CAT GAC AAA GAC AAA GAG TGC AGT TGT AGG GAG TCT GGT TAC 1533
R A S R S Q R K S Q R Q F L R N Q G T P 531
CGT GCC AGC AGA AGC CAA AGA AAG AGT CAA CGG CAA TTC TTG AGA AAC CAG GGG ACT CCA 1593
K Y K P R F V H T R Q T R S L S V E F E 551
AAG TAC AAG CCC AGA TTT GTC CAT ACT CGG CAG ACA CGT TCC TTG TCC GTC GAA TTT GAA 1653
G E I Y D I N L E E E E E L Q V L Q P R 571
GGT GAA ATA TAT GAC ATA AAT CTG GAA GAA GAA GAA GAA TTG CAA GTG TTG CAA CCA AGA 1713
N I A K R H D E G H K G P R D L Q A S S 591
AAC ATT GCT AAG CGT CAT GAT GAA GGC CAC AAG GGG CCA AGA GAT CTC CAG GCT TCC AGT 1773
G G N R G R M L A D S S N A V G P P T T 611
GGT GGC AAC AGG GGC AGG ATG CTG GCA GAT AGC AGC AAC GCC GTG GGC CCA CCT ACC ACT 1833
V R V T H K C F I L P N D S I H C E R E 631
GTC CGA GTG ACA CAC AAG TGT TTT ATT CTT CCC AAT GAC TCT ATC CAT TGT GAG AGA GAA 1893
L Y Q S A R A V K D H K A Y I D K E I E 651
CTG TAC CAA TCG GCC AGA GCG TGG AAG GAC CAT AAG GCA TAC ATT GAC AAA GAG ATT GAA 1953
A L Q D K I K N L R E V R G H L K R R K 671
GCT CTG CAA GAT AAA ATT AAG AAT TTA AGA GAA GTG AGA GGA CAT CTG AAG AGA AGG AAG 2013
P E E C S C S K Q S Y Y N K E K G V K K 291
CCT GAG GAA TGT AGC TGC AGT AAA CAA AGC TAT TAC AAT AAA GAG AAA GGT GTA AAA AAG 2073
Q E K L K S H L H P F K E A A Q E V D S 711
CAA GAG AAA TTA AAG AGC CAT CTT CAC CCA TTC AAG GAG GCT GCT CAG GAA GTA GAT AGC 2133
```

8/44

## FIG. 5C.

K L Q L F K E N N R R R K K E R K E K R 731  
AAA CTG CAA CTT TTC AAG GAG AAC AAC CGT AGG AGG AAG AAG GAG AGG AAG GAG AAG AGA 2193  
R Q R K G E E C S L P G L T C F T H D N 751  
CGG CAG AGG AAG GGG GAA GAG TGC AGC CTG CCT GGC CTC ACT TGC TTC ACG CAT GAC AAC 2253  
N H W Q T A P F W N L G S F C A C T S S 771  
AAC CAC TGG CAG ACA GCC CCG TTC TGG AAC CTG GGA TCT TTC TGT GCT TGC ACG AGT TCT 2313  
N N N T Y W C L R T V N E T H N F L F C 791  
AAC AAT AAC ACC TAC TGG TGT TTG CGT ACA GTT AAT GAG ACG CAT AAT TTT CTT TTC TGT 2373  
E F A T G F L E Y F D M N T D P Y Q L T 811  
GAG TTT GCT ACT GGC TTT TTG GAG TAT TTT GAT ATG AAT ACA GAT CCT TAT CAG CTC ACA 2433  
N T V H T V E R G I L N Q L H V Q L M E 831  
AAT ACA GTG CAC ACG GTA GAA CGA GGC ATT TTG AAT CAG CTA CAC GTA CAA CTA ATG GAG 2493  
L R S C Q G Y K Q C N P R P K N L D V G 851  
CTC AGA AGC TGT CAA GGA TAT AAG CAG TGC AAC CCA AGA CCT AAG AAT CTT GAT GTT GGA 2553  
N K D G G S Y D L H R G Q L W D G V E G 871  
AAT AAA GAT GGA GGA AGC TAT GAC CTA CAC AGA GGA CAG TTA TGG GAT GGA TGG GAA GGT 2613  
\* 872  
TAA 2616

TCAGCCCCGTCCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTAC  
AGACAAAACCTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAAGA  
GTCCTATGAGCAAAATAAAACAAATAAGACTCAAAGTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGC  
TGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTG  
CCAGCTGACCTTCAAACCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAACTTGAATGGAATAACGACATTC  
CAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGA  
AACCGATTTCACTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCAGCCCCAGGCTGCAGCCCATTCGAGGCACCCG  
AAAGAACTTCCCAAGTATGGTGGTCTGGAAAGGACATTTTGAAGATCAACTATATCTTCTGTCATTCCGATGGAA  
TTTCAGTTCATCAGATGTTACCATGGCCACCGCAGAACACCGAAGTAATTCAGCATAGCGGGGAAGATGTTGACCAA  
GGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCTCTGATTAGATG  
AAACTGTTACCCTTACCTAACACAGTATTTCTTTTAACTTTTTTATTTGTAACTAATAAAGGKAATCACAGCCACC  
AACATTTCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACGGAGACT  
CATCGTTATAATTTACTATCTGCCAAGGAGTAGAAAGAAAGGCTGGGGATATTTGGGTGGCTTTGGKTTTGATTTTTT  
GCTTGGTTGGTTGGTTTGKACTAAACAGTATTATCTTTTGAATATCGTAGGGACATAARKWWWMMWKTVWTCMAV  
YMRKAKGSYVRRWVKKGGSTYTYTSKKRKSTMVAMVYKWSMCCYSKKRWVWYVWMMYVCMYKYTSSSTGRYKRN  
KTAATGAAGTT



9/44

Analysis of 23553 (871 aa)

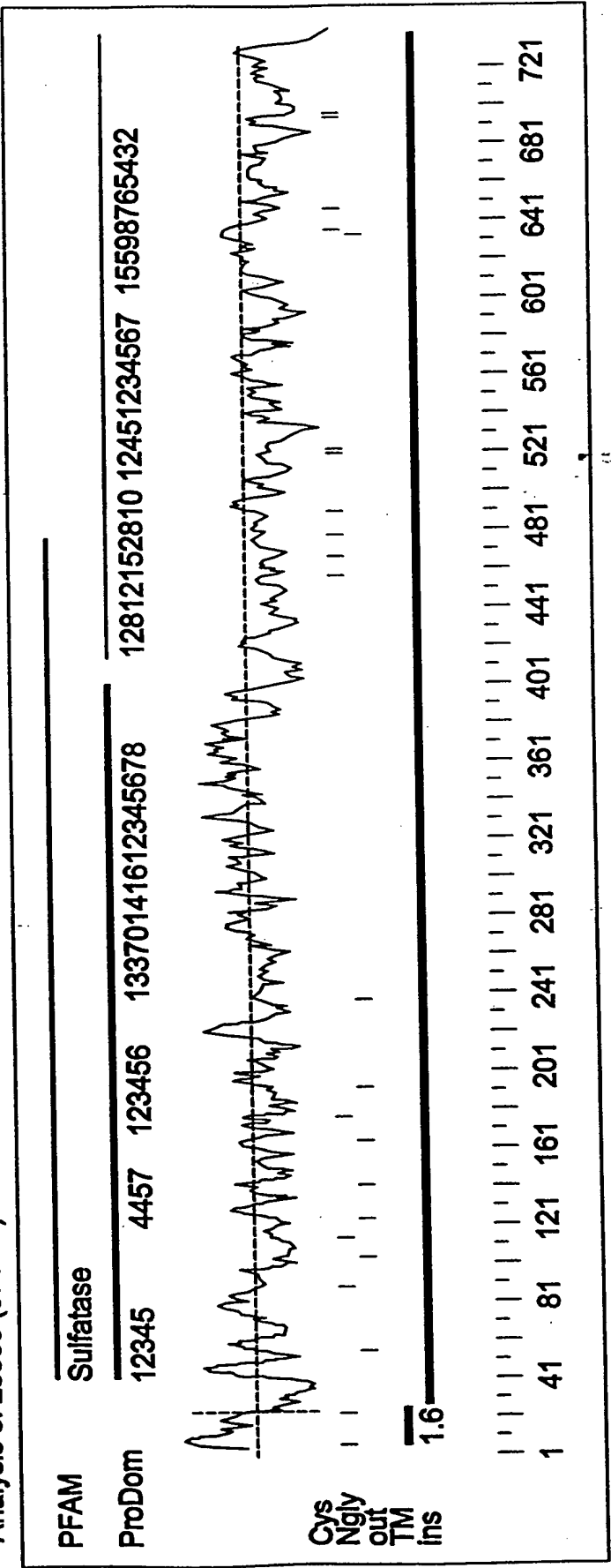
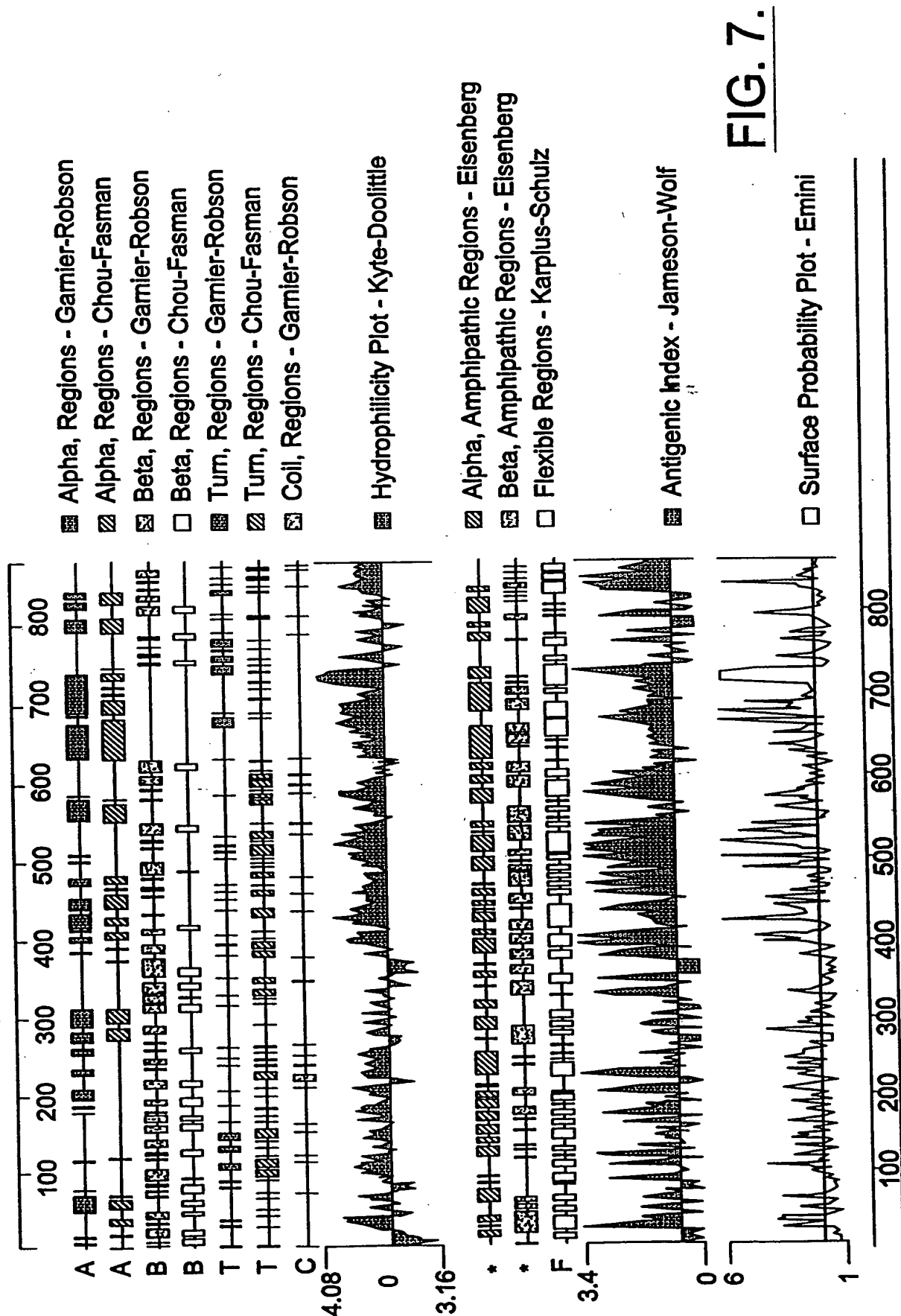


FIG. 6.

10/44



11/44

**Prosite Pattern Matches for 23553**

Prosite versions: Release 12.2 of February 1995

>PS00001/PDOC00001/ASN\_GLYCOSYLATION N-glycosylation site.

Query: 64	NKTR	67
Query: 111	NCSS	114
Query: 131	NNTG	134
Query: 148	NGSY	151
Query: 170	NYTV	173
Query: 197	NESI	200
Query: 240	NASQ	243
Query: 623	NDSI	626
Query: 773	NNTY	776
Query: 783	NETH	786

>PS00005/PDOC00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 24	TVR	26
Query: 27	SPR	29
Query: 66	TRK	68
Query: 96	TGK	98
Query: 206	SKR	208
Query: 400	TNK	402
Query: 425	SSK	427
Query: 468	SGK	470
Query: 484	TVR	486
Query: 488	STR	490
Query: 505	SCR	507
Query: 516	SQR	518
Query: 520	SQR	522
Query: 530	TPK	532
Query: 611	TVR	613
Query: 615	THK	617
Query: 635	SAR	637

>PS00006/PDOC00006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 107	TNNE	110
Query: 288	SVDO	291
Query: 367	TILD	370
Query: 376	TPPD	379
Query: 452	TACE	455
Query: 505	SCRE	508
Query: 781	TVNE	784

**FIG. 8A.**

12/44

>PS00007/PDOC00007/TYR\_PHOSPHO\_SITE Tyrosine kinase phosphorylation site.

Query: 637 RAWKDHKAY 645

>PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 19 GSLCST 24

Query: 161 GLIKNS 166

Query: 325 GLVKGK 330

Query: 592 GGNRGR 597

Query: 763 GSFCAC 768

Query: 851 GNKDGG 856

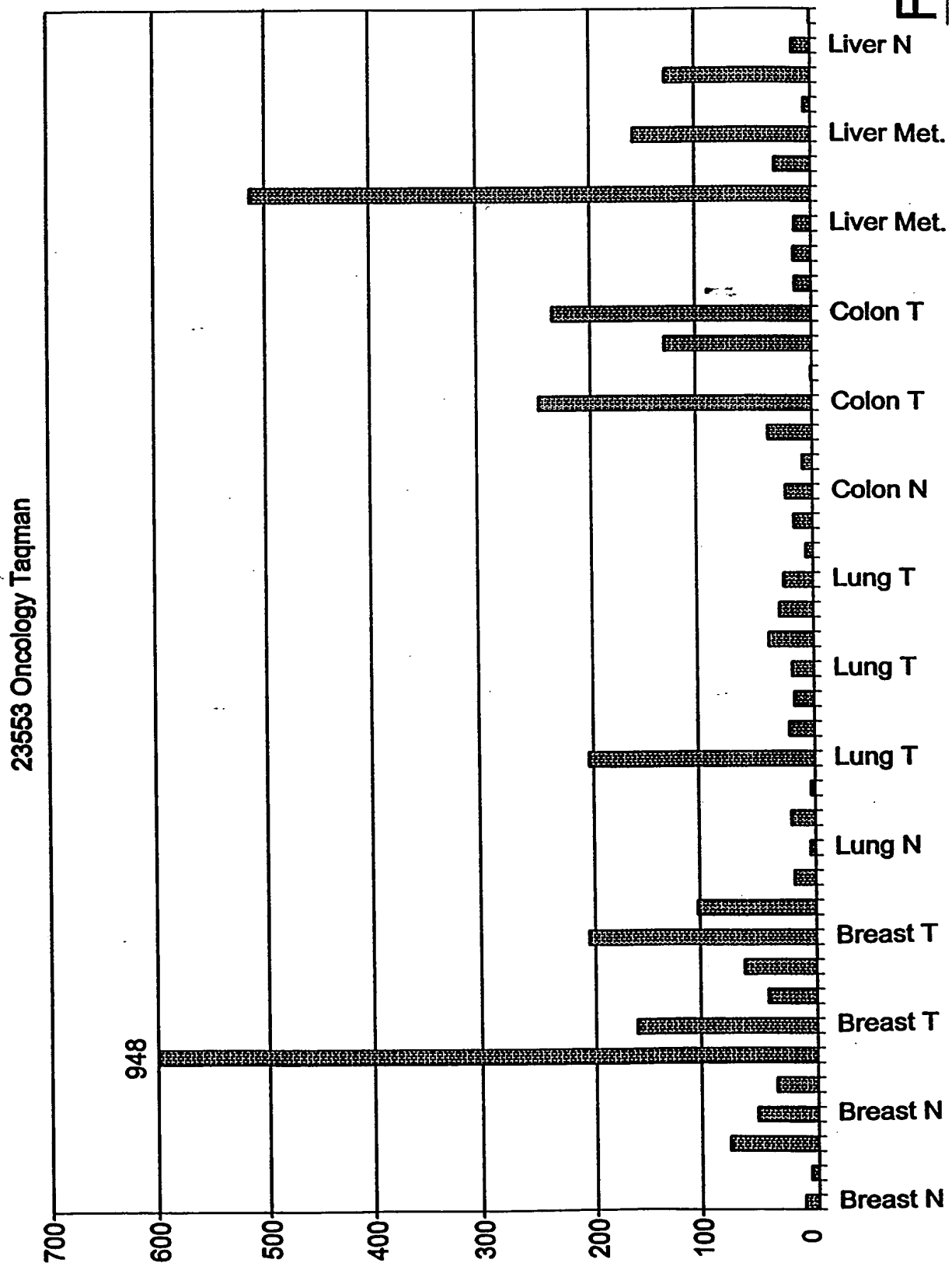
>PS00523/PDOC00117/SULFATASE\_1 Sulfatases signature 1.

Query: 85 PMCCPSRSSMLTG 97

FIG. 8B.

13/44

FIG. 9.



14/44

Input file Fbh25278FL1.seq Output File 25278.trans  
 Sequence length 2940

**FIG. 10A.**

```

CCACGCGTCCGCCCACGCGTCCGGCTGCCACGCCGCGTCTCAGGCTGGCCGGGCTGAGCCGGGAAGAGGGAGCAAAGG
CGGCGCAGGGCCTGCGCTTAGGCAGCGGGAGGCAGCTCGGCGCGGGCCTGACCTCCCCAGAGCGCCCCGCTGCGGGCCGA
GCAGATCCGGCCCAGCCGTCCGGCAGCCAGTCCCGGACCAGACACTGGACCGTCCCCGGGGGCGCTGAACTCCTCGC
AGCATCCGAGCCGGCGGGCCGGTGGTGCGCCCTGGGCGCGCGAGGTGGTGAGGCCCCAGGAGCCCGGCGCGCCGGGACA

      M H T L T G F S L V S L L S F      15
CGCGGGCCGGCTTGGCG ATG CAC ACC CTC ACT GGC TTC TCT CTG GTC AGC CTG CTC AGC TTC      45

  G Y L S W D W A K P S F V A D G P G E A      35
GGC TAC CTG TCC TGG GAC TGG GCC AAG CCG AGC TTC GTG GCC GAC GGG CCC GGG GAG GCT      105

  G E Q P S A A P P Q P P H I I F I L T D      55
GGC GAG CAG CCC TCG GCC GCT CCG CCC CAG CCT CCC CAC ATC ATC TTC ATC CTC ACG GAC      165

  D Q G Y H D V G Y H G S D I E T P T L D      75
GAC CAA GGC TAC CAC GAC GTG GGC TAC CAT GGT TCA GAT ATC GAG ACC CCT ACG CTG GAC      225

  R L A A K G V K L E N Y Y I Q P I C T P      95
AGG CTG GCG GCC AAG GGG GTC AAG TTG GAG AAT TAT TAC ATC CAG CCC ATC TGC ACG CCT      285

  S R S Q L L T G R Y Q I H T G L Q H S I      115
TCG CGG AGC CAG CTC CTC ACT GGC AGG TAC CAG ATC CAC ACA GGA CTC CAG CAT TCC ATC      345

  I R P Q Q P N C L P L D Q V T L P Q K L      135
ATC CGC CCA CAG CAG CCC AAC TGC CTG CCC CTG GAC CAG GTG ACA CTG CCA CAG AAG CTG      405

  Q E A G Y S T H M V G K W H L G F Y R K      155
CAG GAG GCA GGT TAT TCC ACC CAT ATG GTG GGC AAG TGG CAC CTG GGC TTC TAC CGG AAG      465

  E C L P T R R G F D T F L G S L T G N V      175
GAG TGT CTG CCC ACC CGT CGG GGC TTC GAC ACC TTC CTG GGC TCG CTC ACG GGC AAT GTG      525

  D Y Y T Y D N C D G P G V C G F D L H E      195
GAC TAT TAC ACC TAT GAC AAC TGT GAT GGC CCA GGC GTG TGC GGC TTC GAC CTG CAC GAG      585

  G E N V A W G L S G Q Y S T M L Y A Q R      215
GGT GAG AAT GTG GCC TGG GGG CTC AGC GGC CAG TAC TCC ACT ATG CTT TAC GCC CAG CGC      645

  A S H I L A S H S P Q R P L F L Y V A F      235
GCC AGC CAT ATC CTG GCC AGC CAC AGC CCT CAG CGT CCC CTC TTC CTC TAT GTG GCC TTC      705

  Q A V H T P L Q S P R E Y L Y R Y R T M      255
CAG GCA GTA CAC ACA CCC CTG CAG TCC CCT CGT GAG TAC CTG TAC CGC TAC CGC ACC ATG      765

  G N V A R R K Y A A M V T C M D E A V R      275
GGC AAT GTG GCC CGG CGG AAG TAC GCG GCC ATG GTG ACC TGC ATG GAT GAG GCT GTG CGC      825

  N I T W A L K R Y G F Y N N S V I I F S      295
AAC ATC ACC TGG GCC CTC AAG CGC TAC GGT TTC TAC AAC AAC AGT GTC ATC ATC TTC TCC      885

  S D N G G Q T F S G G S N W P L R G R K      315
AGT GAC AAT GGT GGC CAG ACT TTC TCG GGG GGC AGC AAC TGG CCG CTC CGA GGA CGC AAG      945

  G T Y W E G G V R G L G F V H S P L L K      335
GGC ACT TAT TGG GAA GGT GGC GTG CGG GGC CTA GGC TTT GTC CAC AGT CCC CTG CTC AAG      1005

```

15/44

R	K	Q	R	T	S	R	A	L	M	H	I	T	D	W	Y	P	T	L	V	355
CCA	AAG	CAA	CGG	ACA	AGC	CGG	GCA	CTG	ATG	CAC	ATC	ACT	GAC	TGG	TAC	CCG	ACC	CTG	GTG	1065
G	L	A	G	G	T	T	S	A	A	D	G	L	D	G	Y	D	V	W	P	375
GGT	CTG	GCA	GGT	GGT	ACC	ACC	TCA	GCA	GCC	GAT	GGG	CTA	GAT	GGC	TAC	GAC	GTG	TGG	CCG	1125
A	I	S	E	G	R	A	S	P	R	T	E	I	L	H	N	I	D	P	L	395
GCC	ATC	AGC	GAG	GGC	CGG	GCC	TCA	CCA	CGC	ACG	GAG	ATC	CTG	CAC	AAC	ATT	GAC	CCA	CTC	1185
Y	N	H	A	Q	H	G	S	L	E	G	G	F	G	I	W	N	T	A	V	415
TAC	AAC	CAT	GCC	CAG	CAT	GGC	TCC	CTG	GAG	GGC	GGC	TTT	GGC	ATC	TGG	AAC	ACC	GCC	GTG	1245
Q	A	A	I	R	V	G	E	W	K	L	L	T	G	D	P	G	Y	G	D	435
CAG	GCT	GCC	ATC	CGC	GTG	GGT	GAG	TGG	AAG	CTG	CTG	ACA	GGA	GAC	CCC	GCC	TAT	GGC	GAT	1305
V	I	P	P	Q	T	L	A	T	F	P	G	S	W	W	N	L	E	R	M	455
TGG	ATC	CCA	CCG	CAG	ACA	CTG	GCC	ACC	TTC	CCG	GGT	AGC	TGG	TGG	AAC	CTG	GAA	CGA	ATG	1365
A	S	V	R	Q	A	V	W	L	F	N	I	S	A	D	P	Y	E	R	E	475
GCC	AGT	GTC	CGC	CAG	GCC	GTG	TGG	CTC	TTC	AAC	ATC	AGT	GCT	GAC	CCT	TAT	GAA	CGG	GAG	1425
D	L	A	G	Q	R	P	D	V	V	R	T	L	L	A	R	L	A	E	Y	495
GAC	CTG	GCT	GGC	CAG	CGG	CCT	GAT	GTG	GTC	CGC	ACC	CTG	CTG	GCT	CGC	CTG	GCC	GAA	TAT	1485
N	R	T	A	I	P	V	R	Y	P	A	E	N	P	R	A	H	P	D	F	515
AAC	CGC	ACA	GCC	ATC	CCG	GTA	CGC	TAC	CCA	GCT	GAG	AAC	CCC	CGG	GCT	CAT	CCT	GAC	TTT	1545
N	G	G	A	W	G	P	W	A	S	D	E	E	E	E	E	E	E	G	R	535
AAT	GGG	GGT	GCT	TGG	GGG	CCC	TGG	GCC	AGT	GAT	GAG	GAA	GAG	GAG	GAA	GAG	GAA	GGG	AGG	1605
A	R	S	F	S	R	G	R	R	K	K	K	C	K	I	C	K	L	R	S	555
GCT	CGA	AGC	TTC	TCC	CGG	GGT	CGT	CGC	AAG	AAA	AAA	TGC	AAG	ATT	TGC	AAG	CTT	CGA	TCC	1665
F	F	R	K	L	N	T	R	L	M	S	Q	R	I	*						570
TTT	TTC	CGT	AAA	CTC	AAC	ACC	AGG	CTA	ATG	TCC	CAA	CGG	ATC	TGA						1710

TGGTGGGGAGGGAGAAAACGTCTTTAGAGGATCTTCCCCACTCCGGCTTGGCCCTGCTGTTTCTCAGGGAGAAGCCT  
GTCACATCTCCATCTACAGGGAGTTGGAGGGTGAGAGTCCCTTGGTTGAACAGGGTAGGGAGCCTGGATAGGAGTGGG  
TGGGAATAAACAGACTGGGATGCCTGTGTCTCAGTCTGCCTCCTCAGGACTTGCTCTGTGACCTCAGGTGACCCAC  
ATGAGCTTTTAGCCTCAGTTTCTCATCTGTAAATGAGCTCTAATGACTTTGTGACTCTTTGGTGTGGCCCTGGAGCC  
TGGGGCCACGGTGGAGTTCTGGCCGGCCTTGCCACTTGACAACTCCTTTAAGGCTTCCCCCTTAACACGGGATCCCTG  
TGGTGGTGTGTTGGGAGTTGCCTGGAGGCAACTCCAAGCCTGGCCCCAGCTGAAGCATGGCAATCTGGCTGCTCTAC  
AGGGACCCCCAAGCGCTGTGGGTGGAGGGCAGGGGTGGGGGGGTGACCTTCTTGGGTCTTCACATGGCCTAGGCCAG  
TCCTCCGGTCAGACTGGTGTGAGGCACCGTGGTGCAAAATTCCTCTTCTGGCCCTCCAGTACCCAGAGAACTGGCTG  
GGCCATTAACGTGTGAGCACCAAGGGTGGTAGAAAGAGCTGTGAAGAGCCCCCAACAGTACCAGGACACCTGGGTT  
CTCCTGTGACCTGGGGCACAGTTCTTGCCCTCTAGGCCTTGATTCCCCACCTGCAAGTGGGGATGCCAGCCCTGGCTC  
TGCCTCCTTCATGAGGCTCTGGAAGACTGGCCAAGGTTGTGGAGGAGCTTGTAAGTATTAAAGTGTGTAACATGG  
AAAAAAAAAAAAAAAAAAAAAGGGCGG

**FIG. 10B.**

Analysis of 25278 (569 aa)

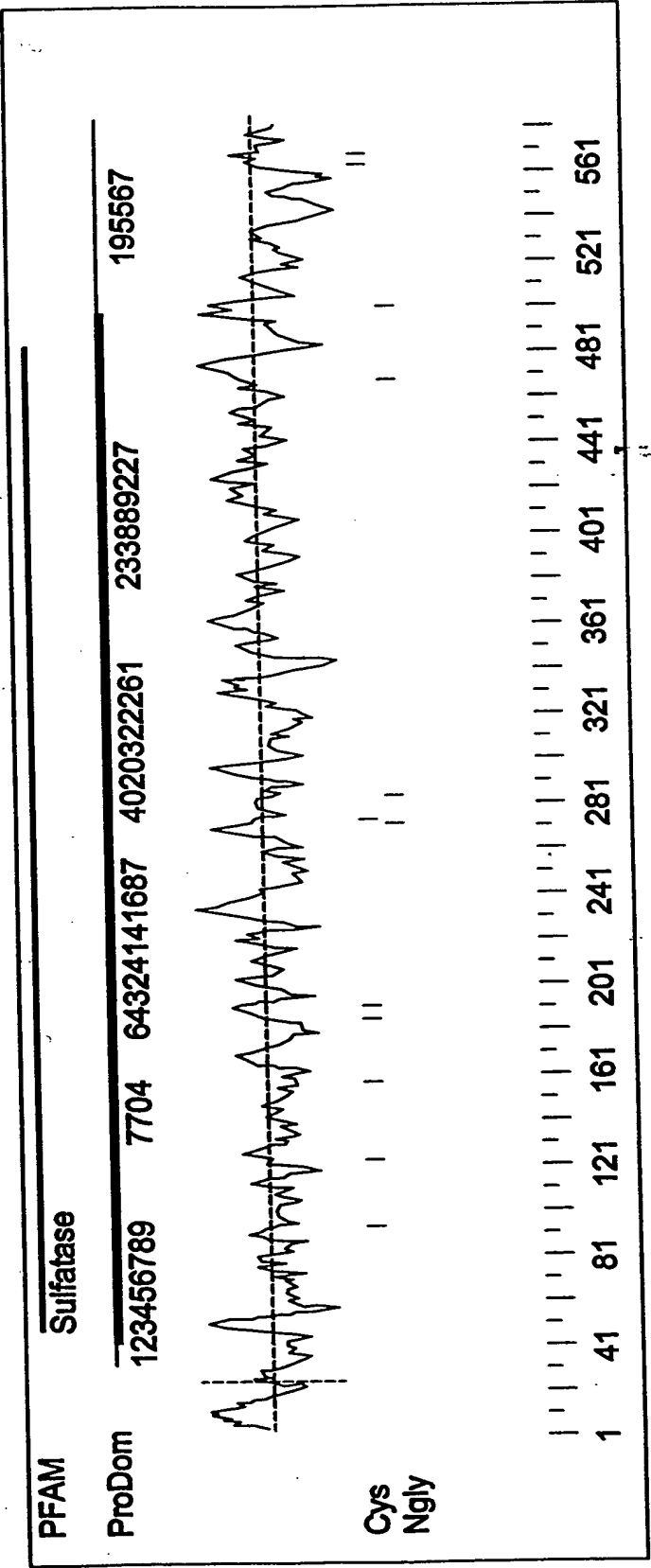
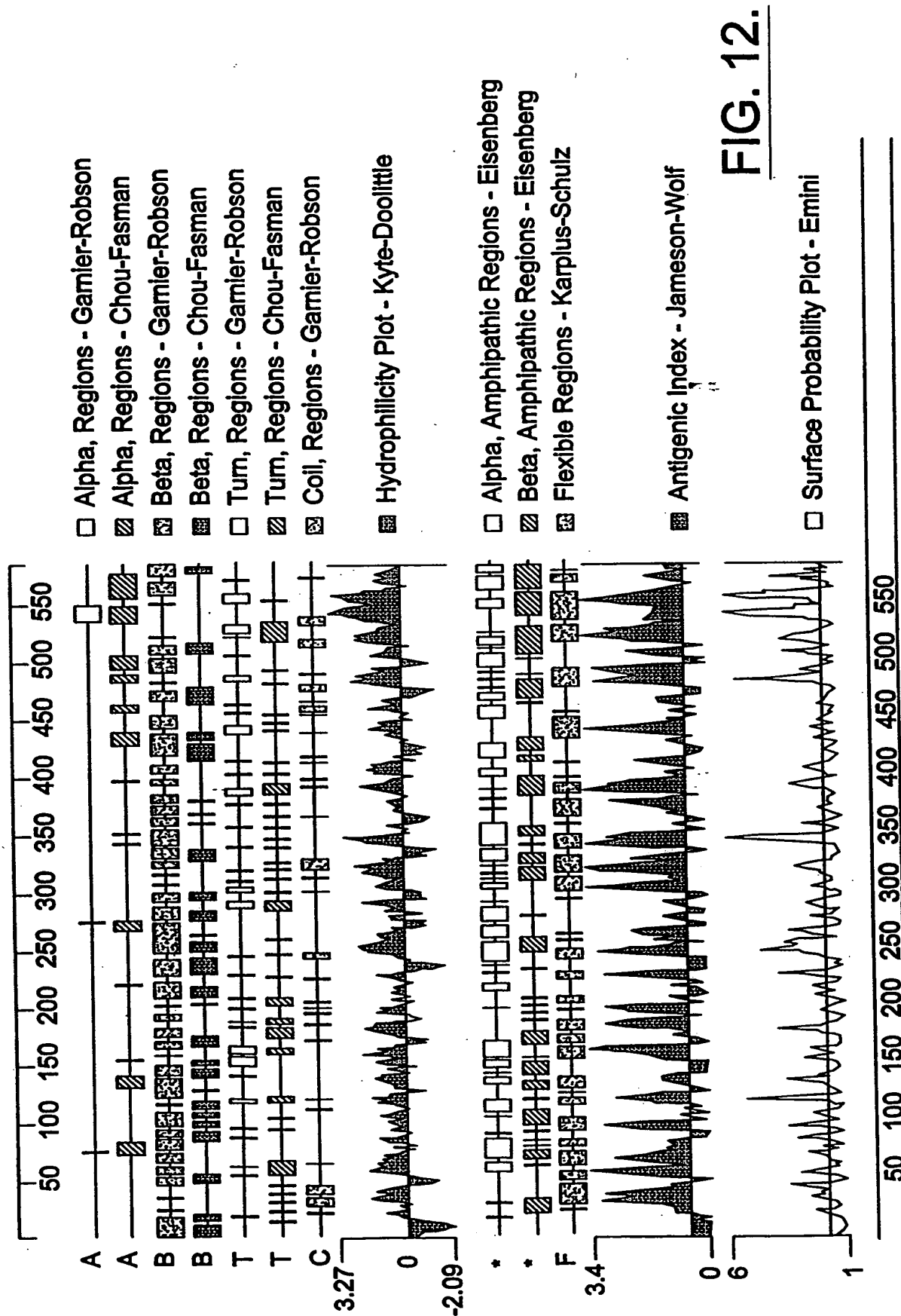


FIG. 11.





## FIG. 13.

### Prosite Pattern Matches for 25278

Prosite versions: Release 12.2 of February 1995

>PS00001/PDOC00001/ASN\_GLYCOSYLATION N-glycosylation site.

Query: 276 NITW 279

Query: 288 NNSV 291

Query: 466 NISA 469

Query: 496 NRTA 499

>PS00004/PDOC00004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 314 RKGK 317

>PS00005/PDOC00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 102 TGR 104

Query: 160 TRR 162

Query: 244 SPR 246

Query: 340 TSR 342

Query: 383 SPR 385

Query: 457 SVR 459

Query: 566 SQR 568

>PS00006/PDOC00006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 67 SDIE 70

Query: 244 SPRE 247

Query: 268 TCMD 271

Query: 317 TYWE 320

Query: 363 SAAD 366

Query: 525 SDEE 528

>PS00007/PDOC00007/TYR\_PHOSPHO\_SITE Tyrosine kinase phosphorylation site.

Query: 134 KLQEAGY 140

>PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 110 GLQHSI 115

Query: 169 GSLTGN 174

Query: 205 GQYSTM 210

Query: 300 GQTFSG 305

Query: 321 GGVRL 326

Query: 356 GLAGGT 361

Query: 402 GSLEGG 407

Query: 409 GWNTA 414

Query: 447 GSWWNL 452

>PS00009/PDOC00009/AMIDATION Amidation site.

Query: 312 RGRK 315

Query: 541 RGRR 544

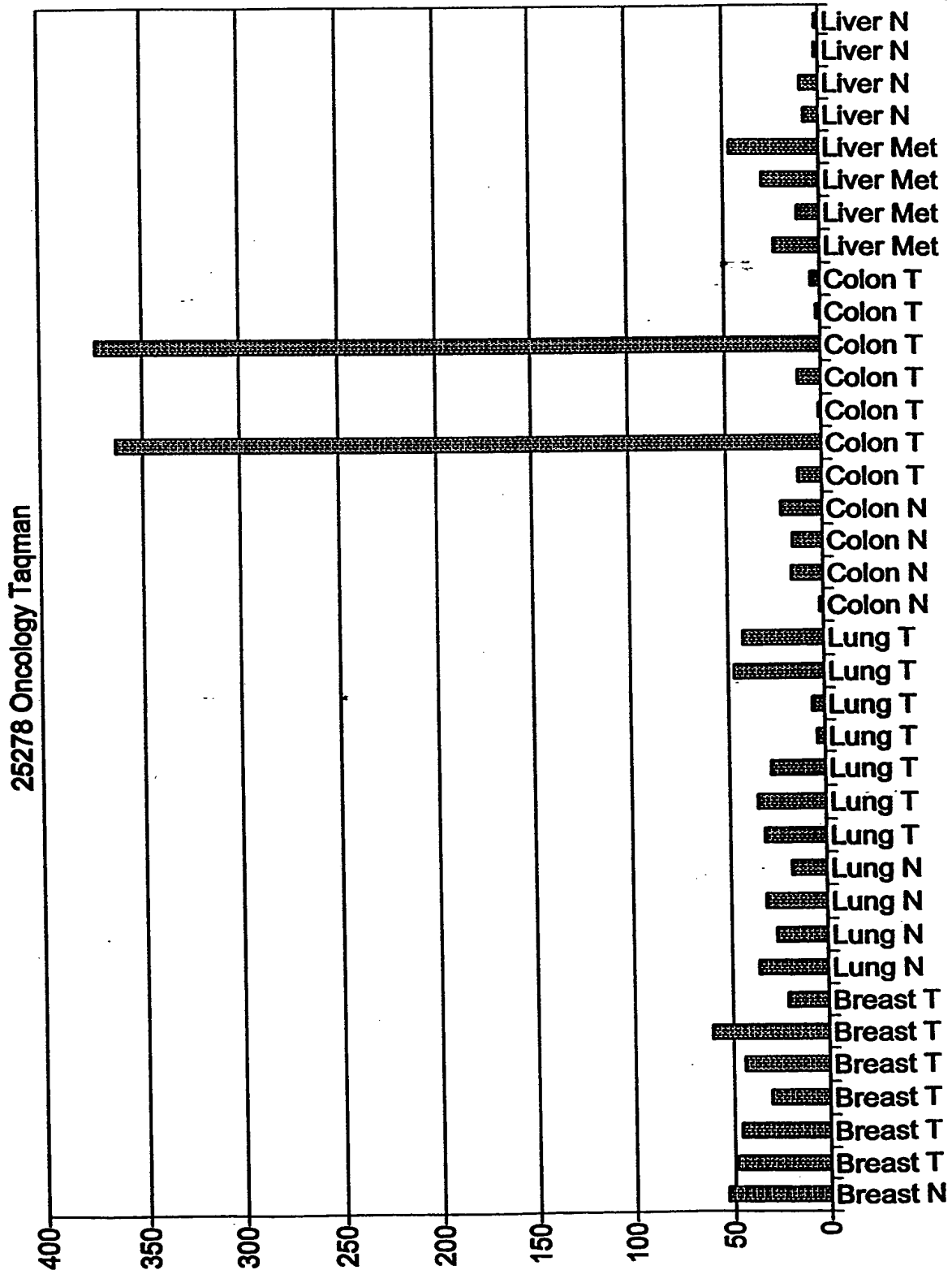
>PS00149/PDOC00117/SULFATASE\_2 Sulfatases signature 2.

Query: 139 GYSTHMVGKW 148

>PS00523/PDOC00117/SULFATASE\_1 Sulfatases signature 1.

Query: 91 PICTPSRSQLLTG 103

FIG. 14.



20/44

26212 seqs  
DNA Sequence (nt 706-2118 coding)

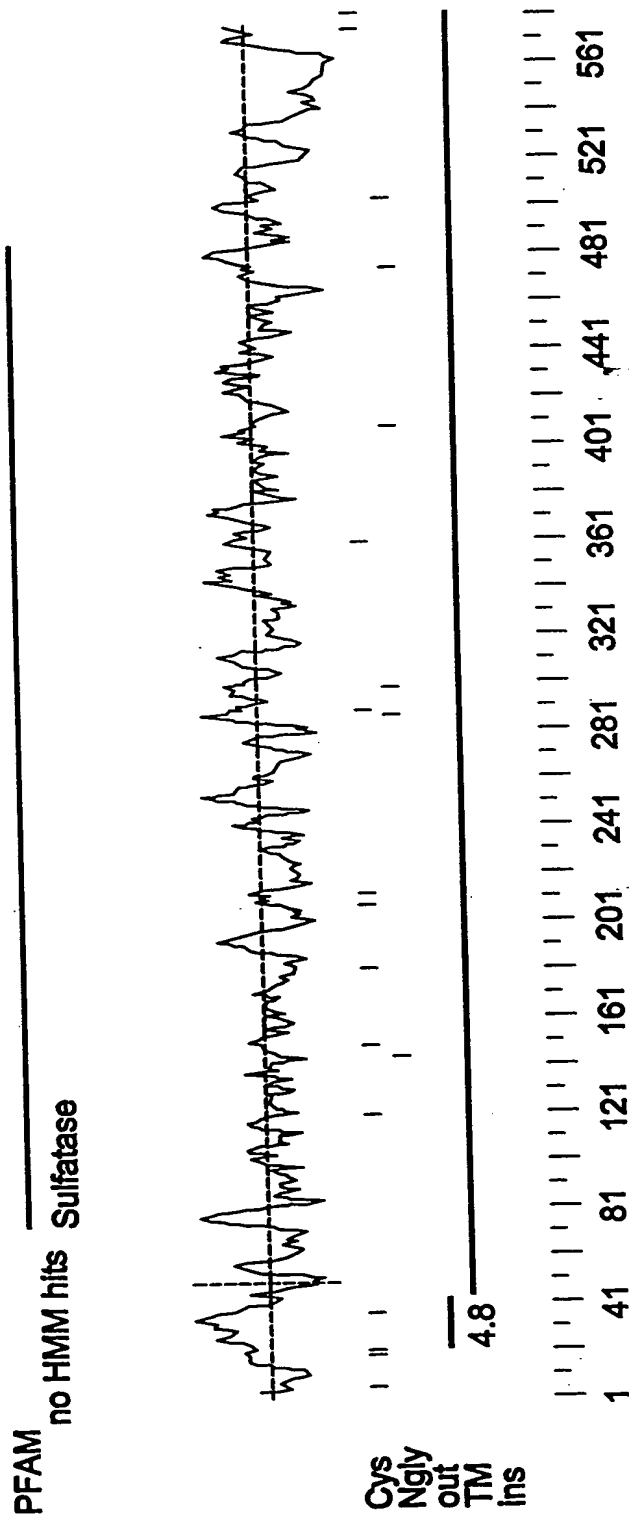
CACGGCTCCGCCACGGCTCCGTGGAGATTAACCTTTTTCTTTTTTTTCTTGGTGAAGCTGCTCTAGGGAGGGGGGAGGAGGA  
GGAGAAAGTGAAATGTGCTGGAGAAGAGCGAGCCCTCTTGTCTTCCGGAGTCCCATCCATTAAGCCATCACTTCTGGAAGATTAAAGT  
TGTCGGACATGGTGACAGCTGAGAGGAGAGGAGGATTTCTTGCCAGGTGGAGAGTCTTACCCTCTGTTGGGTGCATGTGTGCGCCCGCA  
GCGGCGCGGGGCGGTGGTCTCCGCGTGGAGTCTCACCTGGGACCTGAGTGAATGGCTCCAGGGGCTGTGCGGGGCATCCGCCTCCGC  
CTTCTCCACAGGCTGTGTCTGTCTGCTGGAAGATGCTAGCAATGGGGGCGCTGGCAGGATTCTGGATCCTCTGCCTCCTCACTTATGGTT  
ACCTGTCTTGGGGCCAGGCTTAGAAGAGGAGGAAGAAGGGGCTTACTAGCTCAAGCTGGAGAGAACTAGAGCCAGCACAACTTCCA  
CCTCCAGCCCCATCTCATTTTCATCCTAGCGGATGATCAGGGATTAGAGATGTGGTTACCACGGATCTGAGATTAACACCTACTC  
TTGACAGCTCGCTGCCGAAGGAGTTAACTGGAGAACTACTATGTCCAGCTATTTGCACACCATCCAGGAGTCAGTTTATTACTGGA  
AGTATCAGATACACACCGGACTTCAACATTCTATCATAAGACCTACCAACCAACTGTTTACCTCTGGACAATGCCACCTACCTCAGA  
AACTGAAGGAGGTGGATATTCACGCATATGGTCGGAATGGCACTTGGGTTTTACAGAAAAGAAATGCATGCCACCAGAGAGGAT  
TTGATACCTTTTTTGGTCCCTTTTGGGAAGTGGGGATTACTATACACTACAAATGTGACAGTCTGGGATGTGTGGCTATGACTTGT  
ATGAAAACGACAATGCTGCGTGGGACTATGACAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAATCTTAGCTTCCC  
ATAACCCACAAAGCTATATTTTATATATTGCTATCAAGCTGTTCACTTCAACTGCAAGCTCCTGGCAGGTATTTGAACTACTC  
GATCCATTATCAACATAACAGGAGGAGATATGCTGCCATGCTTCTGCTTAGATGAAGCAATCAACAACGTGACATTGGCTCTAAAGA  
CTTATGGTTTCTATAACAACAGCATTATCATTTACTCTTCAGATAATGGTGGCAGCCTACGGCAGGAGGGAGTAAGTGGCTCTCAGAG  
GTAGCAAGGAACATATTGGGAAGGAGGATCCGGGCTGTAGGCTTGTGTCATAGCCACTTCTGAAAAACAAGGGAACAGTGTGTAGG  
AAGTGTGACATCACTGACTGGTACCCCACTCTCATTTCACTGGCTGAAGGACAGATTGATGAGGACATTCACTAGATGGCTATGATA  
TCTGGGAGACCATAAGTGAGGTCTTCGCTACCCCGAGTAGATATTTGCATAACATTGACCCCATATACCAAGGCAAAAAATGGCT  
CCTGGGCAGCAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTGCAGCACTGGAATGCTTACAGGAATCCTGGCT  
ACAGCGACTGGGTCCCCCTCAGTCTTTCAGCAACCTGGGACCGAACCGGTGGCACAATGAACGGATCACCTTGTCACTGGCAAAAGTG  
TATGGCTTTTCAACATCACAGCCGACCATATGAGAGGGTGGACCTATCTAACAGGTATCCAGGAATCGTGAAGAGCTCTACGGAGGC  
TCTCAGTTTCAACAAAATGCAGTGGCGTCAAGTATCCCCCAAGACCCAGAGTAACCTAGGCTCAATGGAGGGGTCTGGGGAC  
CATGGTATAAAGAGGAAACCAAGAAAAAGCAAGCAAAATCAGGCTGAGAAAAAGCAAAAGAAAAAGCAAAAGAAAGAAAC  
AGCAGAAAGCAGTCTCAGGTTCAACTTGCCATTGAGGTGTACTTGTGGATAAGCACAATATTTCTGTGTTGGTTAACTTTAATCAGT  
TCTTATCTTTCATCTGTTTCTAGGTAACAGCAAAATTTGGCTCGATAATATCGCTGGCTAAGCGTCAGGCTTGTTTTCATGCTGTGC  
CAC

Protein sequence

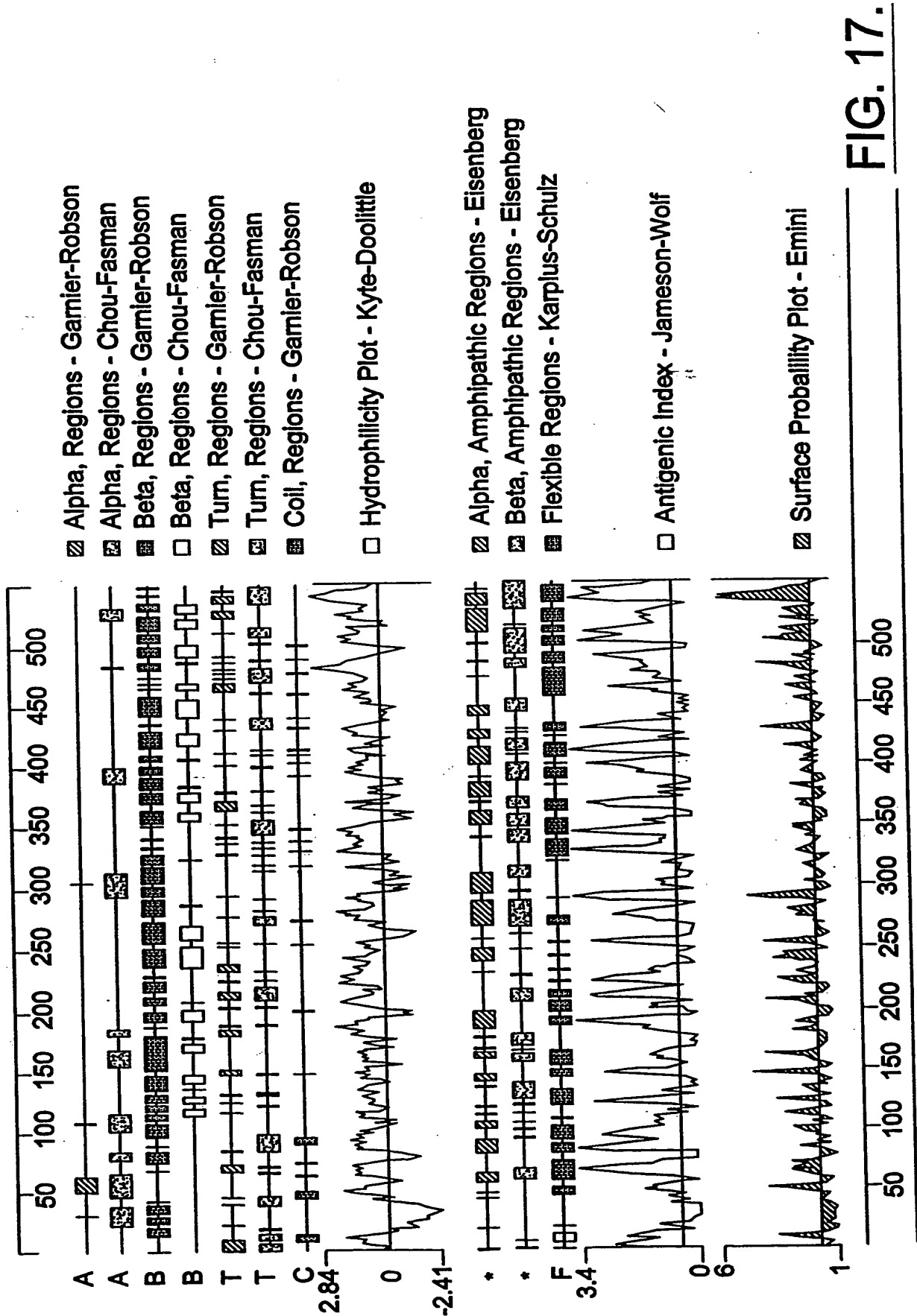
MAPRGCAHPPPPSPQACVCPGKMLAMGALAGFWILCLLYGYLSVGGQALEEEEEGALLAQAGEKLEPSTTSTSQPHLIFILADDQGRFD  
VGYHGEIKPTLDKLAAGVLENYYVQPICTPSRSQFITGKYQIHTGLQHSIRPTQPNCLPLDNATLPQKLKEVGYSTHMGVHLG  
FYRKECMPTRRGFDITFFGSLGSGDYTHYKCDSPGMCGYDLYENDNAVDYDNGIYSTQHYTQRVQQILASHNPTKPIFLYIAYQAVHS  
PLQAPGRYFEHYRSIININRRRYAAMLSCLEAINNVTALKTYGFYNNSIIYSSDNGGQPTAGGSNVPLRGSKGTVEGGIRAVGFVH  
SPLLKNKGTVCKELVHITDVYPTLISLAEGQIDEDIQLDGYDIVETISEGLRSPVDILHNIDPIYTKAKNGSVAAGYGIWNTATQSAIR  
VQHVKLLTGNPGYSWVPPQSFNLGPNRVHNERITLSTGKSVLFINITADPYERVDLNRYPGIVKLLRRLSQFNKTAVPVRYPPKDP  
RSNRLNGGVGVPWYKEETKKKKPSKNQAEKKQKSKKKKKQKAVSGSTCHSGVTCG

FIG. 15.

**FIG. 16.**



22/44



### Prosite Pattern Matches for 26212

Prosite version: Release 12.2 of February 1995

>PS00001/PDOC00001/ASN\_GLYCOSYLATION N-glycosylation site.

Query: 157 NATL 160  
Query: 306 NVTL 309  
Query: 318 NNSI 321  
Query: 431 NGSW 434  
Query: 497 NITA 500  
Query: 527 NKTA 530

>PS00004/PDOC00004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 521 RRLS 524  
Query: 562 KKPS 565

>PS00005/PDOC00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 131 TGK 133  
Query: 189 TRR 191  
Query: 243 TQR 245  
Query: 413 SPR 415  
Query: 489 TGK 491  
Query: 509 SNR 511  
Query: 559 TKK 561  
Query: 576 SKK 578

>PS00006/PDOC00006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

**FIG. 18A.**

24/44

Query: 298 SCLD 301  
Query: 347 TYWE 350  
Query: 386 SLAE 389  
Query: 406 TISE 409

>PS00007/PDOC00007/TYR\_PHOSPHO\_SITE Tyrosine kinase phosphorylation site.

Query: 163 KLKEVGY 169

>PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 28 GALAGF 33  
Query: 56 GALLAQ 61  
Query: 139 GLQHSI 144  
Query: 198 GSLLGS 203  
Query: 235 GIYSTQ 240  
Query: 329 GGQPTA 334  
Query: 343 GSKGTY 348  
Query: 351 GGIRAV 356  
Query: 432 GSWAAG 437  
Query: 439 GIWNTA 444

>PS00149/PDOC00117/SULFATASE\_2 Sulfatases signature 2.

Query: 168 GYSTHMGKW 177

>PS00523/PDOC00117/SULFATASE\_1 Sulfatases signature 1.

Query: 120 PICTPSRSQFITG 132

**FIG. 18B.**



## 25/44

Alignments of top-scoring domains:

Sulfatase: domain 1 of 1, from 36 to 462: score 323.0, E = 3.5e-93

```

*->PNiIlIlaDDIGidIGcyGnptIrtPniDrLAeeGlrFtnayvttP
PN+++ilaDD+G+gdIG+ + t t n+D +A+eG+rF ++ +++
25277 36 PNFVII LADDMGWGDLGANVAETKDTANLDKMASEGMRVFDFHAAAS 82

1CtPSRAaILLTGRyphrtGmytnnragvlpftgwsleGglpldettlpeI
C+PSRA+ILTGR+ r+G++ n + +s +Gglpl+ettl+e+
25277 83 TCSPSRASLLTGRLGLRNGVTRNF AV-----TS-VGGLPNETTLAEV 124

LkeaGYaTgmVGKVLgyneessasdfahIPlgrGFdyfygnlGGEdQWY
L++aGY+Tg++GKVLg++ ++ +P rGFdy+tg
25277 125 LQQAGYVTGIIGKVLGHGHSY-----HPMFRGF DYFG----- 158

plvdal lpftndtytceggyskdvallkplgalgvneveapdkaladyk
+pt++ ++ c
25277 159 -----IPYSH-DMGCT-----D 169

tagaInvphhvFEVadryagavdvgrpflavli fprpaacf lypnatvvs
t+g+ + p + + + + + tr + ++ + a+ ly nt +v+
25277 170 TPGYNGPP-----CPACPGDGP SRNLQRDCY--TDVALPLYENLNIVE 211

qpmphspltPrpwl ladealpf lerngqrpkpflylsykhvhlprda
qp s l+ q +a++a +f++r+ + +pf+ly++++h+h+p
25277 212 QPVNLSSLA-----QKYAEKATQFIQRASTSGRPFLLYVALAHMHVP--- 253

pnlfsskdfagssrrglYglllDsveemDdgvgrvlnaLdeINGlldnTl
l+ + a r lYg + + emD +vg++ + +d + +nT+
25277 254 --LPVTQLPAAPRGRSLYG---AGLWEMDSL VGQIKDKVDHT--VKENTF 296

iIFTSllDhGghlgahghlgiragGsnpgfrg.....gKgtnllye
FT D+G+ ++ + + Gs gpf g +++++++K+t+ te
25277 297 LVFTG--DNGPWAQKCELA---GSVGPFTGfwqtrqqqspAKQTT-VE 338

gGtRvpIivrWPeGIlapggvsdelvslnDlFPTll dLAGapIPgvaagv
gGtRvP++++wP Gt + + +s +l s++D+fPT+++LA a+IP
25277 339 GGHVRPALAYWP-GRVPVNTSTALLSVLDIFPTVVALAQASLP----- 381

kdrilDGvsLlpLLgaagssrhetlfyesyncnegrgflpavrwgkkkah
+ r DGv++ ++L g+ + +h lf++ n g a++ +
25277 382 QGRRFDGVDVSEVLFGR-SQPGHRVLFHP---NSG-----AAGEFGALQT 422

frtpniagwqrvdlddvkllfntvedfnrsgddacrhgdvckclgkprrs
+r + + k+f++++ + + + g+ + +
25277 423 VRLE-----RYKAFYITGGAR---ACDGTSGPELQHKF----- 452

vthhdppllydlrDP<-*
pl ++l+ D
25277 453 -----PLIFNLEDDT 462
  
```

**FIG. 19.**

26/44

Alignments of top-scoring domains:

Sulfatase: domain 1 of 1, from 43 to 467: score 268.9, E = 6.5e-77

```

*->PNIllllaDDlGlgdlGcyGnptirtpnIDrLAeeGlrFtnayvttp
      PNI+l+l+DD++ ++lG+      ++ ++      + +G F na+vttp
23553   43   PNIILVLTDDQD-VELGSLQ---VMNKTRKIMEHGGAIFAFVTTTP 85

      1CtPSRAaILTGRyphrtGmytnnravglpftgwsleGglpldettlpeI
      +C+PSR++ LTG+y h++++ytnn++      +++++ w+      ++ +++++
23553   86   MCCPSRSSMLTGKYVHNHNVYTNNEN--CSSPSWQ----AMHEPRTFAVY 129

      LkeaGYaTgmVGKWHlgyneessasdfahIPlgrG.FdyfygnlGGEdQW
      L + GY+T+++GK++++yn ++      +P+g+ ++      +n
23553   130  LNNTGYRTAFFGKYLNEYNGSY-----IPPGWReWLGLIKNI----- 165

      Yplvdal lpfndtytcegggygfskdvalkp lga lgvneveapdkalady
      ++f+n + c+tg      + +++++ +++++      dy
23553   166  -----SRFYN-YTVCRNG-----IKEKHGF DYAK-----DY 190

      ktagalnvphhvFEWadryagavdvgrpflavl lfrpaacfl ypnatvv
      +++++n +      y++++      p+++++      +
23553   191  FTDLITNES-----IN YFKMSK-----RMYPHRPVMMV-----I 219

      sqmphpsltaPrpwql ladealpf lerngardkpf flylsykhvhiprd
      s+ +ph p + +      +++++ + p+ + + +      + +++++ +++++
23553   220  SHAAPHGPED-S---APQFSKLYPNASQH-ITPSYNYAPNMDKHWIMQYT 264

      apnlfsskdfagssrrglYgl l lDsveemDdgvgvrlnaLde lNGl ldnT
      +pnl+ + +ft +r++ +      +++++Dd+vr++n L e G+l+nT
23553   265  GPMLPIHMEFTNlLQKRLQ-----TLMSVDDSVRLYNMLVET-GELENT 309

      l l lFTS l lDhGghlgahghlgiragGsnpgfrggKgtnlyegGtRvPliv
      +l l+T+ DHG+h+g++g+      + gK+++ y++++RvP+++
23553   310  YIIYTA--DHGYKIGQFGGLV-----K-GKSMP-YDFDIRVPFFI 344

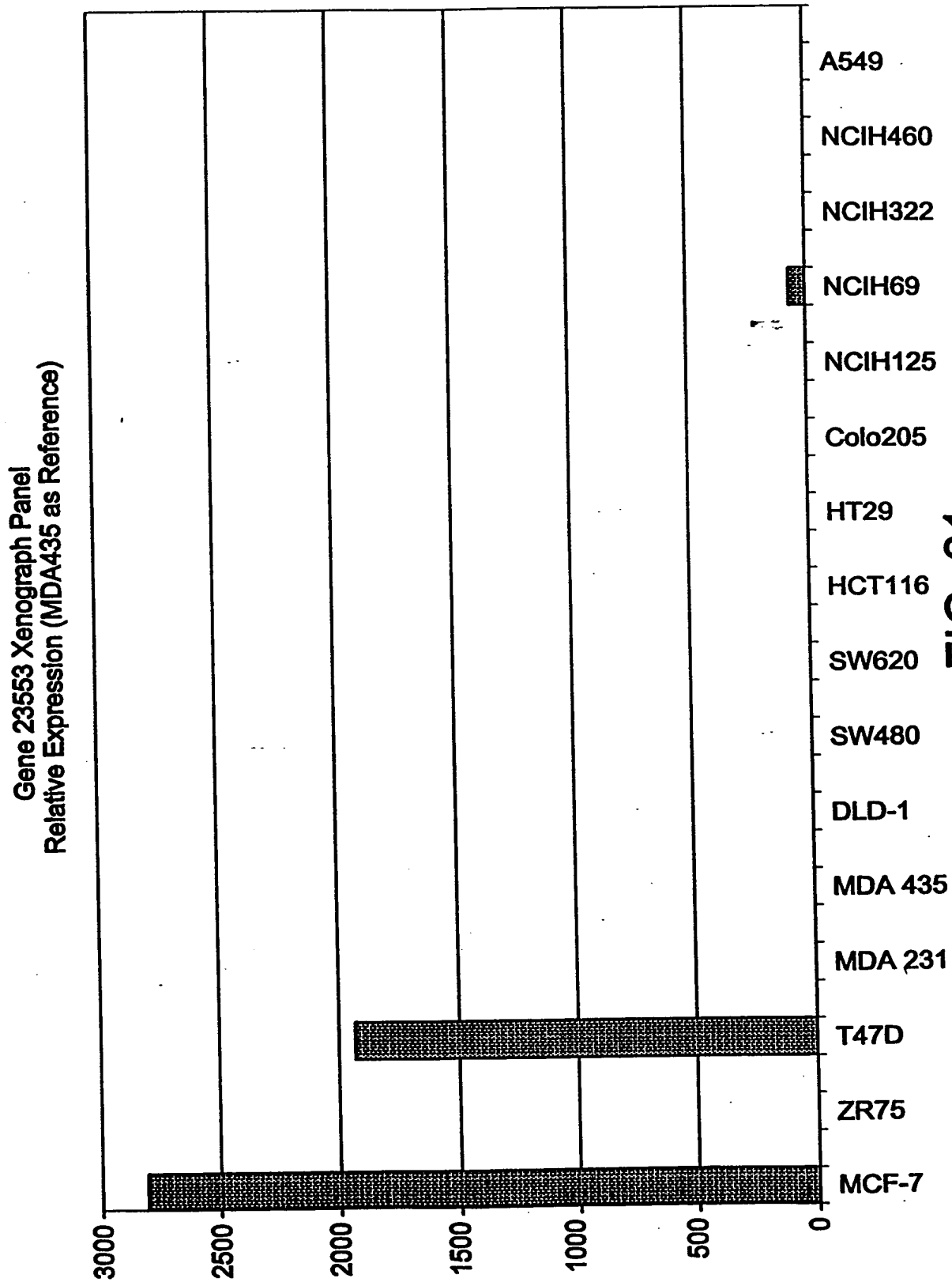
      rwPeG l l apgqvdselvs l mD l fPT l l dLAGap l Pgvaagvkdr l l DGvs
      r+p      +pg+++ ++v ++D l +PT l l d+AG++ P      +DG+s
23553   345  RGP--SVEPGSIVPQIVLNlDlAPTl l DIAGlDTP-----PDVDGKS 384

      L l p l l l gaagssrhet l fyesycnegrgf lpavrwgkkkahfrtpnl agw
      +l+lL+ +      ++ ++f + + +++++ +++++ + +f
23553   385  VLKLLDPE----KPGNRFRT-NKKAK----IWRDTFLVERGKF----- 418

      qrvdfddvwlk l fntvedfnrsgddacrhgdvckclgkprsvthhdppl l
      + k + + + +s++ + + +      +c +++++ ++ +++++
23553   419  -----LRKKEESSKNlQQSNHLPKYERVKELCQQARYQTA-CEQPQGK 460

      ydl srDPK-*
      +D
23553   461  VQCI EDT 467
  
```

**FIG. 20.**



**FIG. 21.**

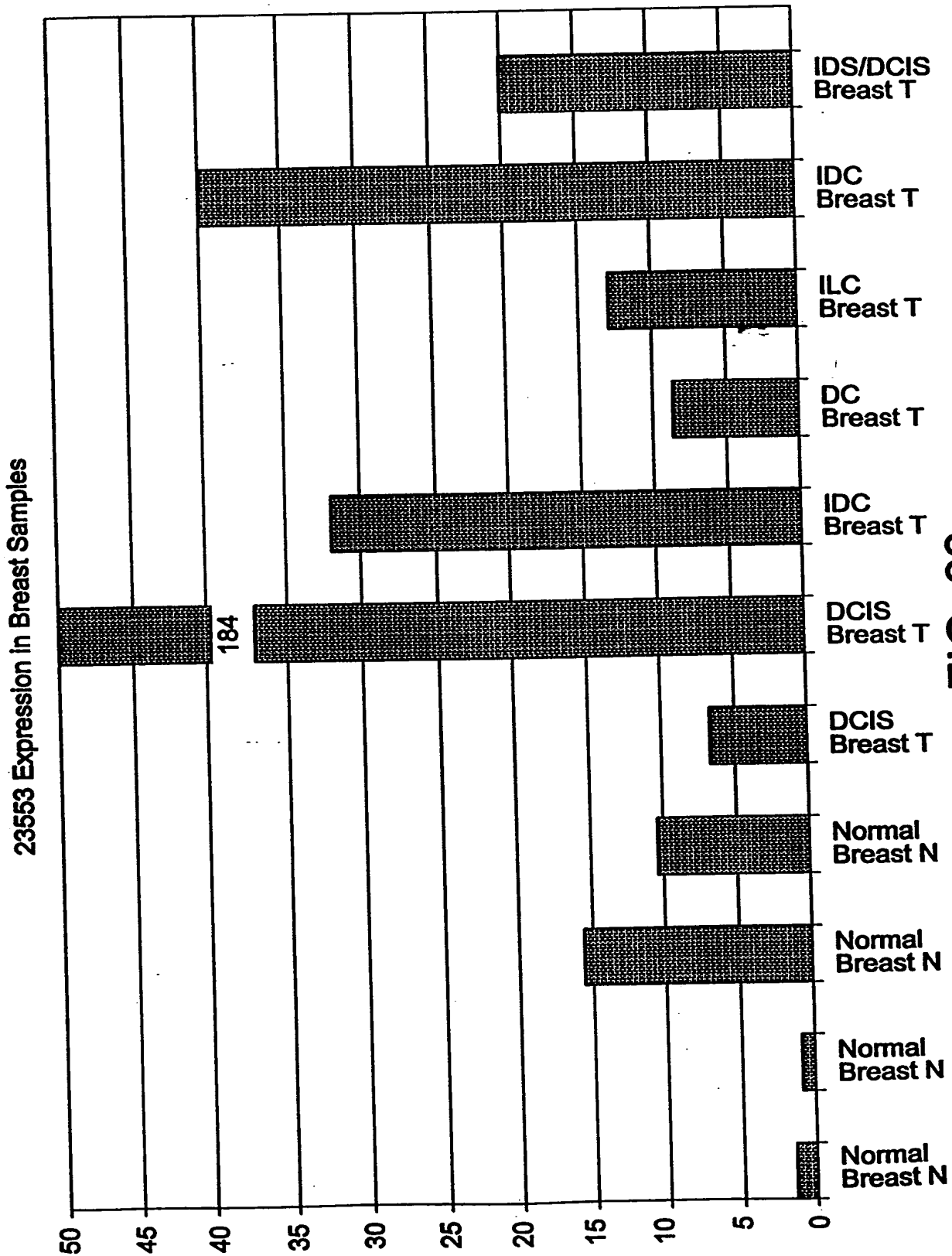


FIG. 22.

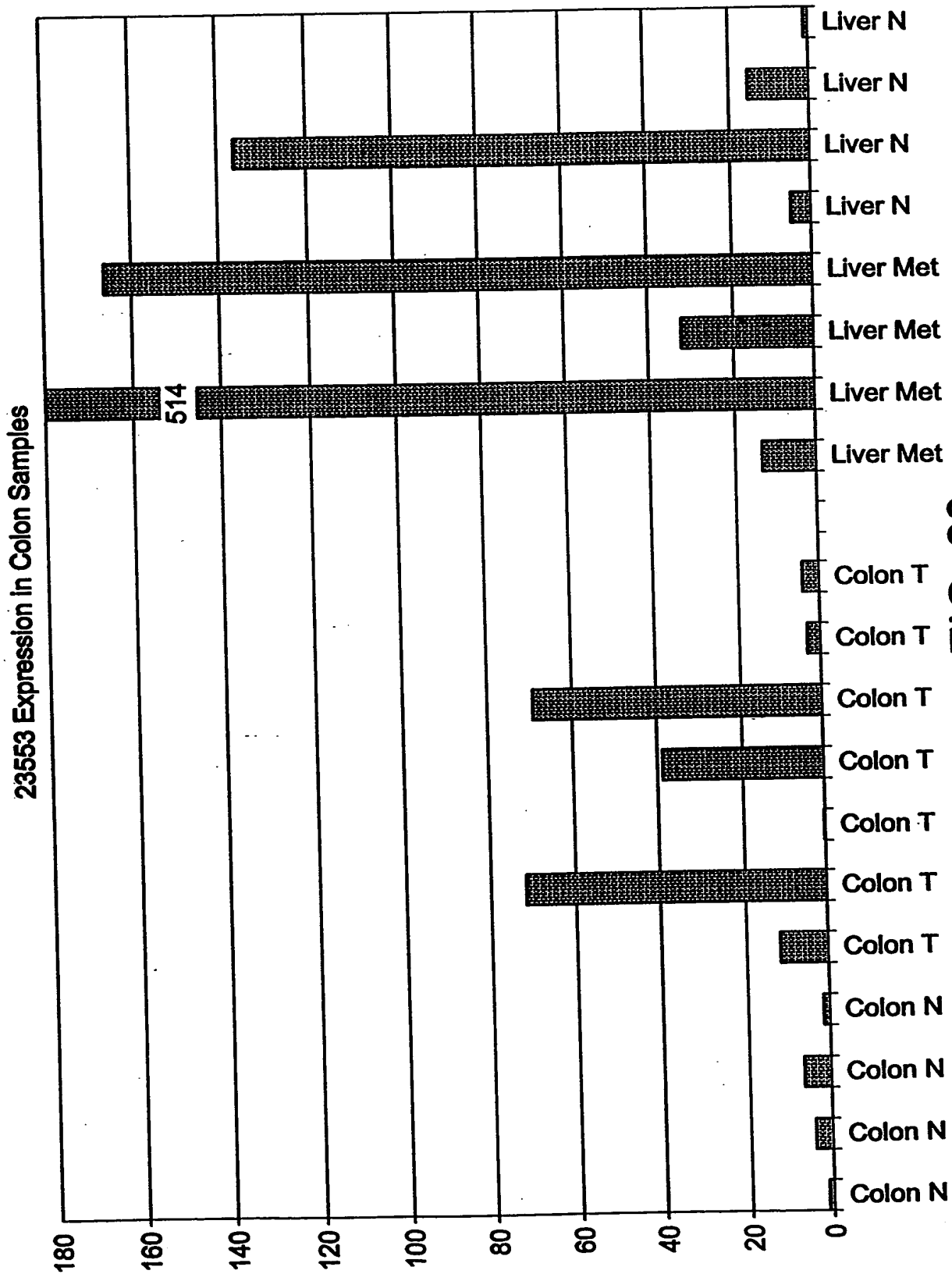


FIG. 23.

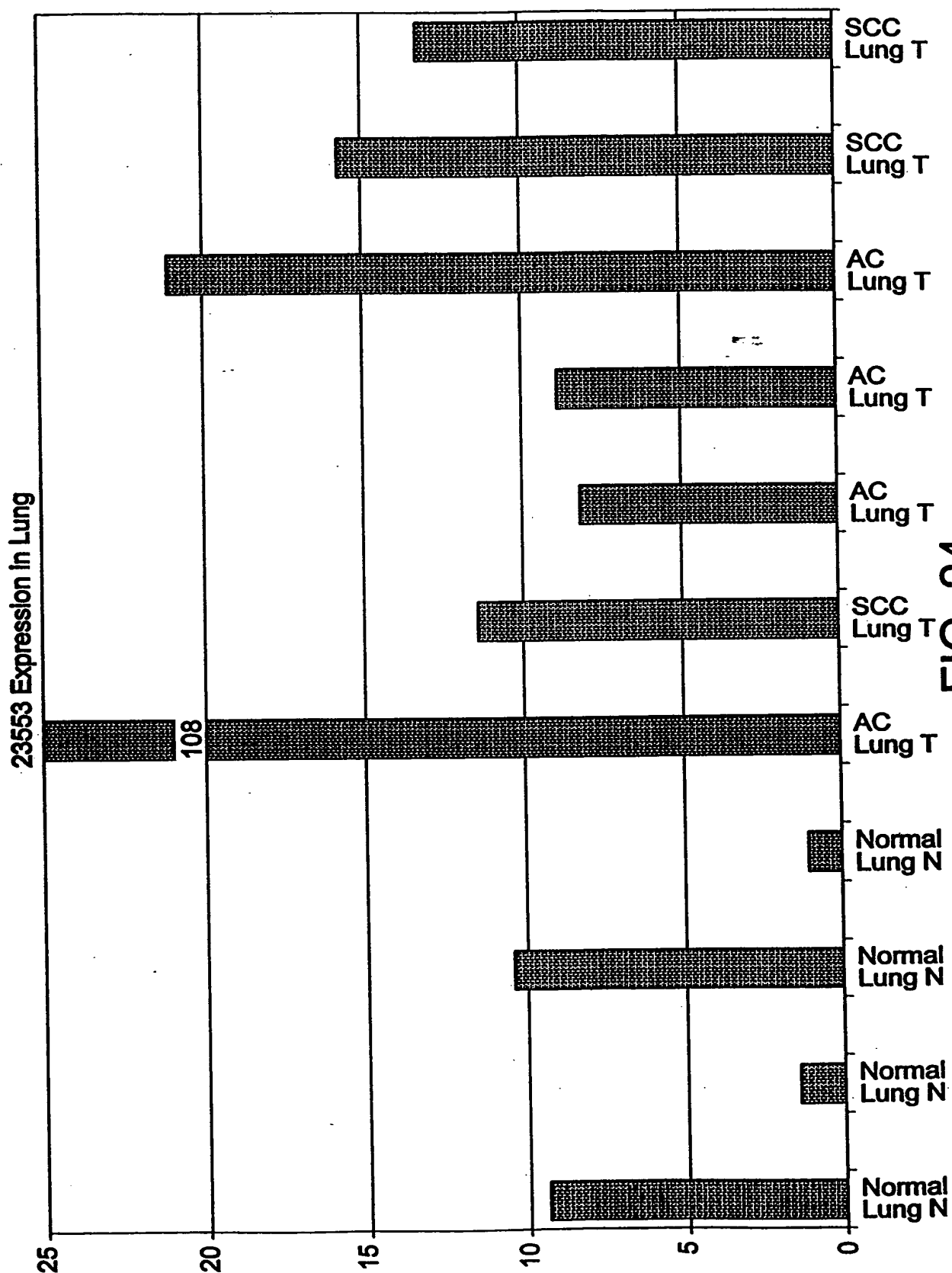


FIG. 24.

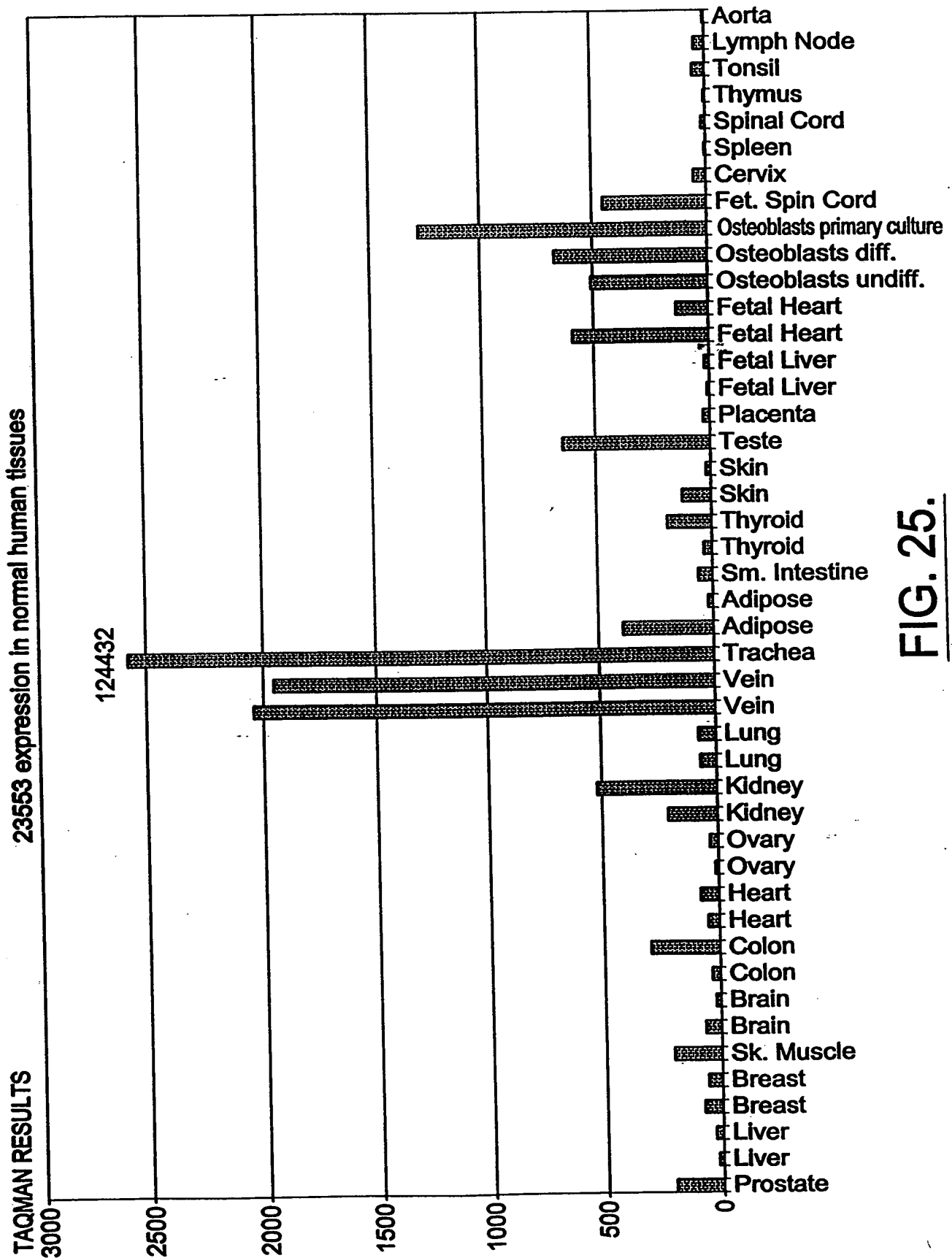
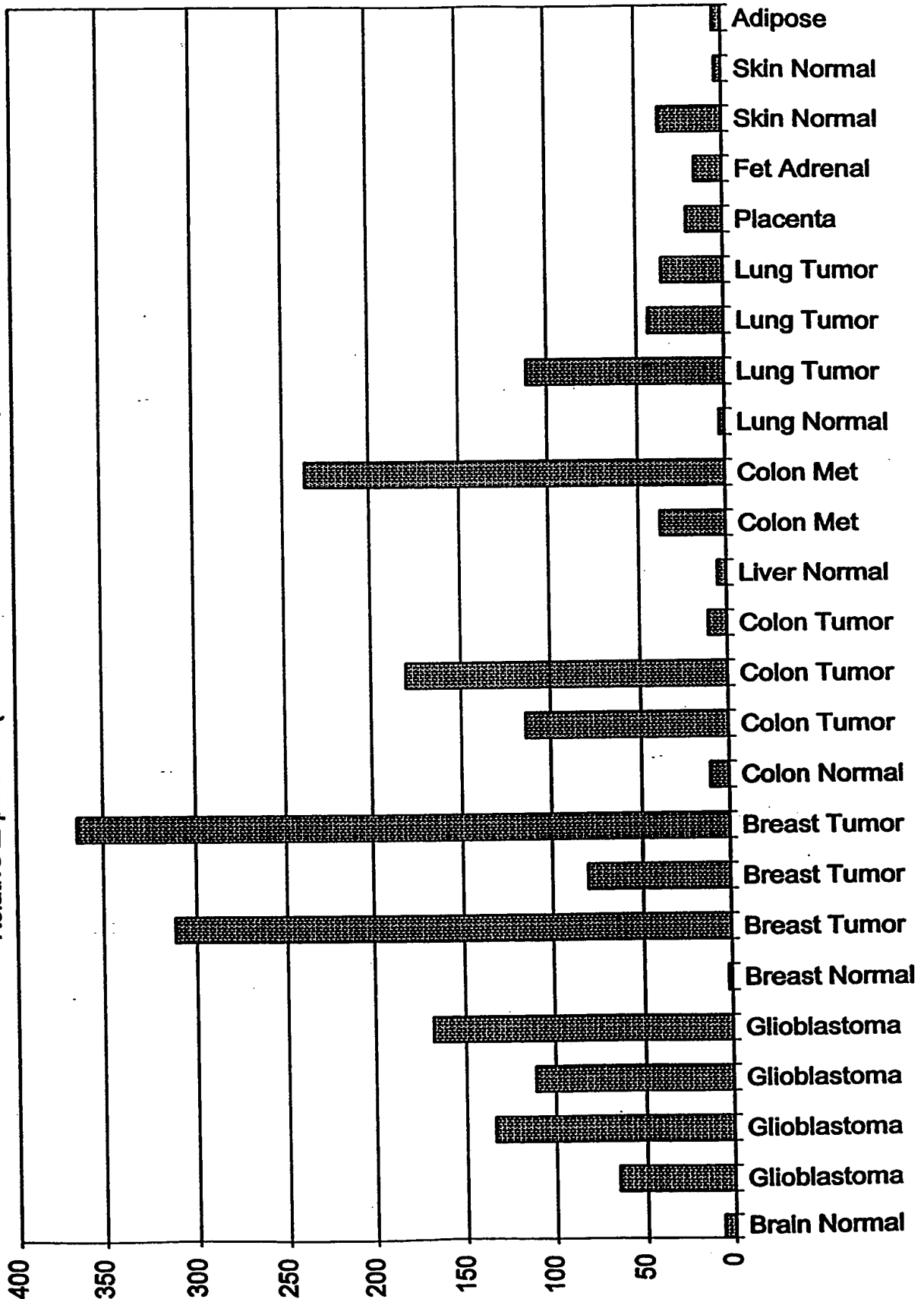


FIG. 25.

**FIG. 26.**

Gene 23553 Oncology Panel  
Relative Expression (Normal Breast as Reference)





33/44

Alignments of top-scoring domains:

Sulfatase: domain 1 of 1, from 47 to 471: score 289.7, E = 3.6e-83

```

*->PNiIlilaDDIGigdlGcyGnptlrtpniDrLAeeGlrFtnayvttp
P+i++i++DD+G+ d+G +G + i+tp++DrLA+ G+++ n y+ tp
25278    47    PHIIFILTDDQGYHDVGYHG-SDIETPTLDRLAAKGVKLEN-YYIQP 91

    lCtPSRAaLTGRyphrtGmytnnragvlpftgwsleGglpldettlpeI
+CtPSR++LTGRy+++tG+++ + p+++ +lpId +tlp+
25278    92    ICTPSRSQLLTGRYQIHTGLQHSIIR---PQQPN---CLPLDQVTLPOK 134

    LkeaGYaTgnvGKWHlglyneessasdfahIPlgrGFdyfygnlGGEdQWY
L+eaGY T+nvGKWHlg +++++ IP++rGFd+f+g+
25278    135   LQEAGYSTHMGVGHVGFYRKEC-----LPTRRGFDITFLGS----- 170

    plvdal lpftndtytcegggygfskdvalkplgalgvneveapdkaladyk
l + d+yt+++ ++
25278    171   -----LTGNVDYYTYDN-----CD 184

    tagalnvphhvfEWadryagavdvgrpf lavli fprpaacflypnatvvs
+g+ + +d + + + +
25278    185   GPGVCG-----FD-----LHEGENVAVG 202

    qpmphspltPrpwql ladealpf lerngqrkpf flylsykhvhiprd.
++s++ +a++a l ++ +p fly+++++vh+p ++
25278    203   LSGQYSTML-----YAQRASHILASH-SPQRPLFLYVAFQAVHTPLQ5 244

    apmIfsskdfagssrrglYglilDsveemDdgvgvrvlnaLdeINGlIdnT
+ +++ ++ g+ r+ Y+ ++v nD++v ++ aL++ G ++n
25278    245   PREYLYRYRTMGNVARRKYA---AMVTCMDEAVRNITWALKRY-GFYNN5 290

    lIIFTSIlDhGghlgahghlgiragGsngpfrggKgtilyegGtRvPliv
+IIF+S D+Gg++ gGsn+p+rg+Kgt +egG+R ++v
25278    291   VIIFSS--DNGGQTF----S----GGSNWPLRGRKGTY-WEQGVRLGFV 329

    rWPeGIlapggysdelvslmDlFPTlIdLAGapIPgvaagvkdrIDGvs
++P +++ ++s++l ++ D++PT++ LAG++ + lDg++
25278    330   HSP-LLKRKQRTSRALMHITDWPYTLVGLAGGTTS-----AADGLDGYD 372

    LlpLLgaagssrhetlfye.....syncnegrgflpavrwgkkkahfrt
++p++ ++ +s+e+l++++ ++ +++ g + g ++ ++
25278    373   VVPAISEGRASPTEILHNIIdplynhaQHGSLEG-----GFGIWNTAVQA 417

    pni. agwqrvdfddvwlfnftvedfnrsgddacrhdvckclgkprsvt
+ + w + + ++d+ +++ a + g + + ++
25278    418   AIRvGEWK-----LLTGDPGYGDWIPPQTLATFPGSWWNLER----MAS 457

    hhdppll ydlSrDP<-*
+ l+++s+DP
25278    458   VRQAVWLFNISADP 471

```

FIG. 27.

**34/44**

			<u>Relative Expression</u>
NDR 19	Breast T	DCIS	47.84
MDA 138	Breast N	Normal	52.89
NDR 01	Breast T	IDC	44.79
NDR 15	Breast T	DC	29.55
NDR 133	Breast T	ILC	43.26
MDA 161	Breast T	IDC	60.13
MDA 155	Breast T	IDC/DCIS	20.11
PIT 270	Lung N	Normal	36.00
CHT 427	Lung N	Normal	26.54
PIT 241	Lung N	Normal	31.45
PIT 298	Lung N	Normal	17.57
CHT 800	Lung T	AC	31.45
CHT 335	Lung T	SCC	35.02
CHT 447	Lung T	AC	27.19
CHT 752	Lung T	AC	3.89
CHT 799	Lung T	AC	5.74
CHT 813	Lung T	SCC	47.18
CHT 369	Lung T	SCC	42.37
CHT 371	Colon N	Normal	2.37
CHT 396	Colon N	Normal	16.34
CHT 398	Colon N	Normal	15.24
NDR 104	Colon N	Normal	20.89
CHT 520	Colon T	Adeno	11.71
CHT 122	Colon T	Adeno	360.79
CHT 536	Colon T	Adeno	1.00

**FIG. 28A.**

**35/44**

			<u>Relative Expression</u>
CHT 528	Colon T	Adeno	11.63
CHT 386	Colon T	Adeno	372.22
CHT 372	Colon T	Adeno	2.39
CHT 532	Colon T	Adeno	4.45
CHT 77	Liver Met	Met	23.43
CHT 321	Liver Met	Met	11.35
CHT 84	Liver Met	Met	30.38
NDR 100	Liver Met	Met	46.21
NDR 154	Liver N	Normal	7.31
CHT 322	Liver N	Normal	9.38
PIT 51	Liver N	Normal	1.77
CHT 339	Liver	Normal	1.58
PIT 265	Breast N	Normal	37.40
MDA 335	Breast N	Normal	45.57
NDR 132	Breast T	DCIS	19.56
NDR 13	Breast N	Normal	6.73
NDR 56	Breast N	Normal	20.61

**FIG. 28B.**

36/44

Alignments of top-scoring domains:

Sulfatase: domain 1 of 1, from 76 to 502: score 324.5, E = 1.3e-93

```

*->PNvllllaDDlGlgdlycyghptlrTPnldrLAeeGlrFtnhytatp
P+ ++llaDD+G+ d+g++g ++l TP+ld+LA+eG+++ n+y+ +p
26212 76 PHLIFILADDQGFDRDVGyHG-SEIKTPTLDKLAAGVKLENYYV-QP 120

lCsPSRAaLlTGryphrhGmvsngrlgvlgftaksgglpldettLpellk
+C+PSR+++ TG+y++++G + + + ++ +lpld +tLp+ Lk
26212 121 ICTPSRSQFITGKYQIHTGLQH-----SIIRPTQPNCLPLDNATLPQKLK 165

eaGYaTglvGKWHlglnensdaagdgehlpGwrGfdyfdgflygspfty
e GY T++vGKWHlg+++ +et P++ rGfd f+g l+gs ++y
26212 166 EVGYSTHVMVGKWHLGfYR-----KECMPTR-RGFDTFGSLGSGDY 207

deencdngegteppeaypeqgwlpqilgyyltdlladkalglldvasaag
++ cd +pt +++++l+ aa
26212 208 THYKCD-----SPGM-----CGYDLYENDNAA- 229

rllakalaasrPFlyispaphfsilfrnfkevaqpyrapqltqlfvde
++++ + +tq++++
26212 230 -----WDYD-----NGIYSTQMYTOR 245

aadfiernk.ekPfflylafirlhvhtplfspaedleskdfigrsqrgrY
+++++ kP fly att +vh pl++p + e+++ r+rY
26212 246 VQQILASHNpTKPIFLYIAYQ--AVHSPLQAPGRYFEHYRSIININRRRY 293

gdIveemDdlvGrvldaLedlGllldNTlvifTSDnGahlegtpewygggn
+++++ D+++++v aL+ G ++N ++++SDnG g+pt +gg+n
26212 294 AAMLSCLDEAINNVTLALKTYGFYNNSIIYSSDNG----GQPT-AGGSN 338

gpIkggKgygslyeGgIRvPlIvrwPggiapagrvekselvshvDlaPT
+pl+g Kg+ +eGgIR ++v++P + +g+v + elv++ D++PT
26212 339 WPLRGSKGTy--VEGGIRAVGFVHSP-LLKNKGTVCk--ELVHITDwyPT 383

lldlAGapIPkvanGakdrplDGvsllpIIlggaapsrrahetlfhyngk
+ +lA + ++ d lDG++++ + +g + s+ + +h+
26212 384 LISLAEGQIDE-----DIQLDGYDIWETISEGLR-SP--RVDILHN--- 421

grklravrwprkskgtptklkahffftpat.....
++ ++ +k+ + + a + ++ ++ + ++ + +++++ ++
26212 422 ___IDPIYTKAKN---GSAAGYGIWNTaIgsaIrvqhwklltgnpgysd 465

.... dddtnngwecvgtvsqaddiedcrcegvettvthdppelyDlsrDP
++++ n+g + ++ e t+ + +l++ ++DP
26212 466 wvppQSFNLG-----PNRWHNER-ITLSTGKSVVLFNITADP 502

<-*
```

26212 - -

FIG. 29.

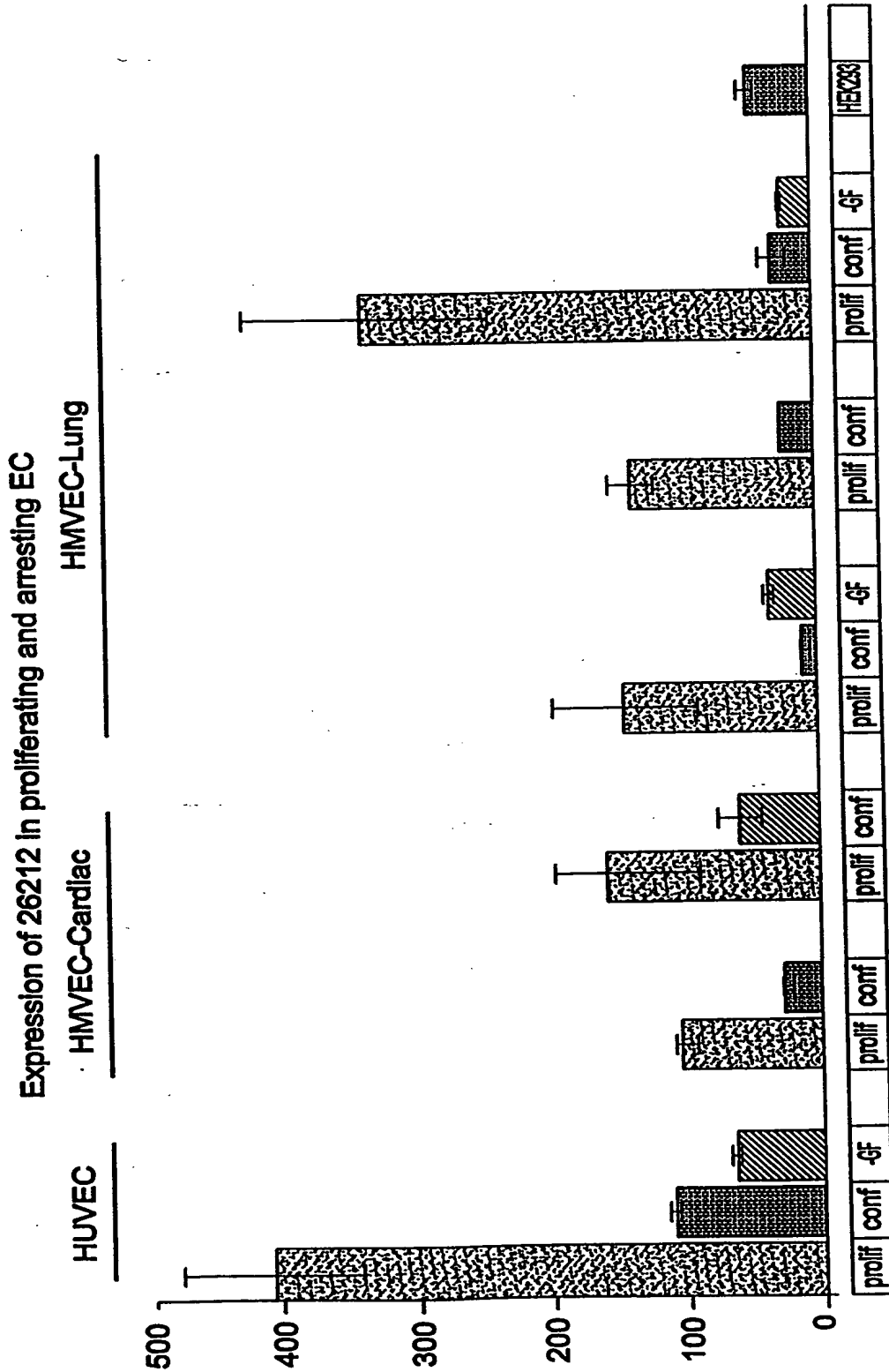


FIG. 30.

38/44

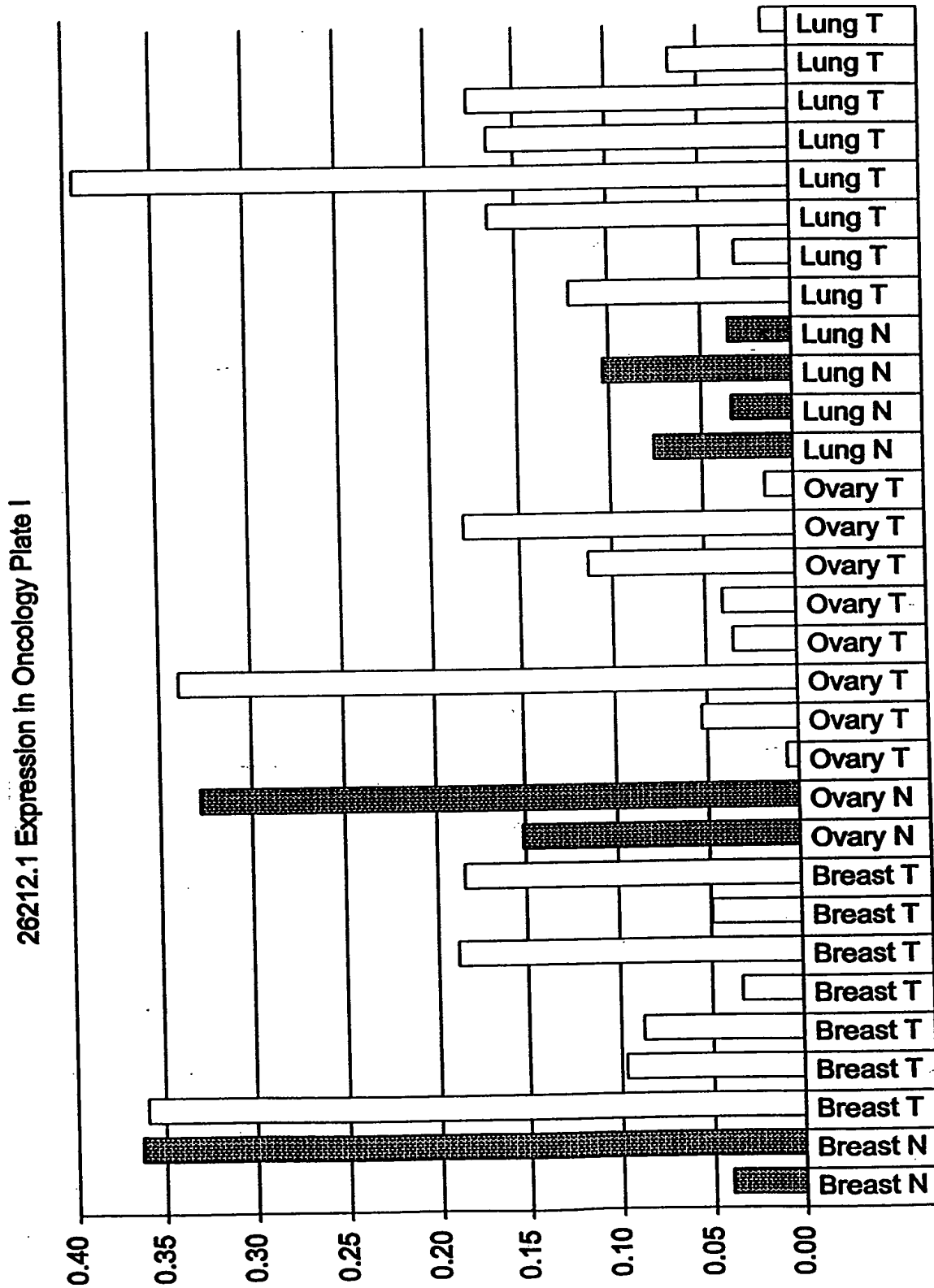
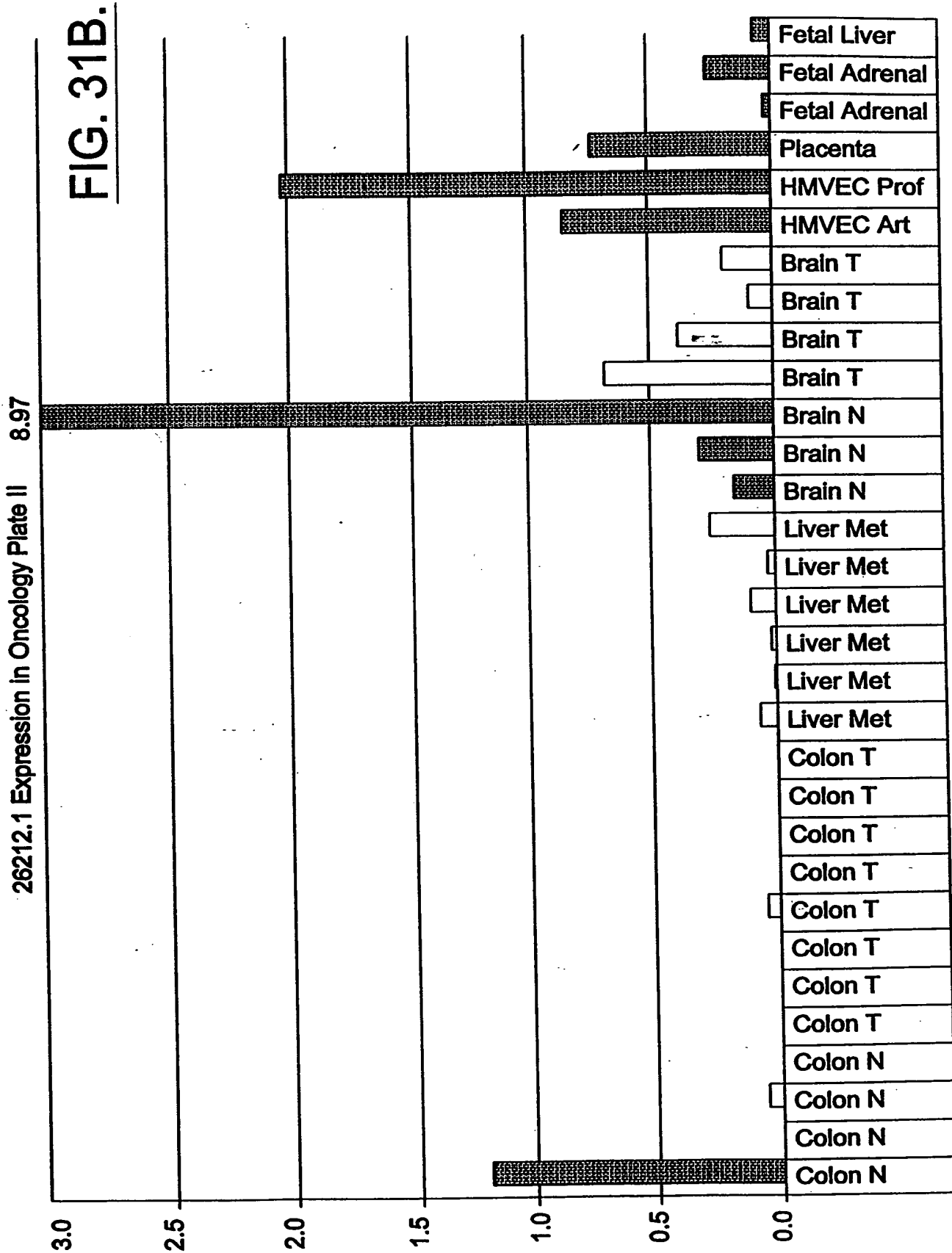
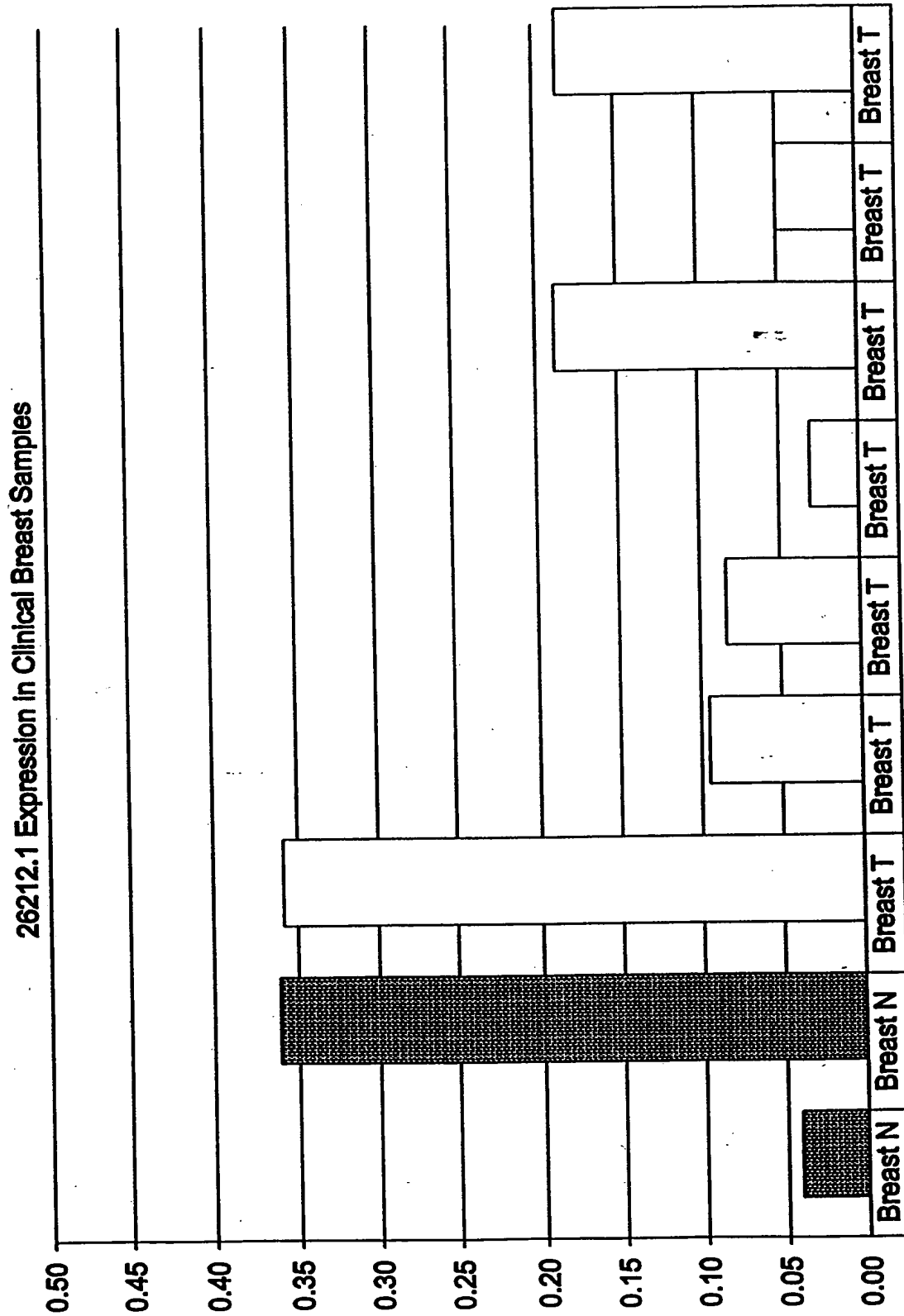


FIG. 31A.





**FIG. 32.**



26212.1 Expression in Clinical Lung Samples

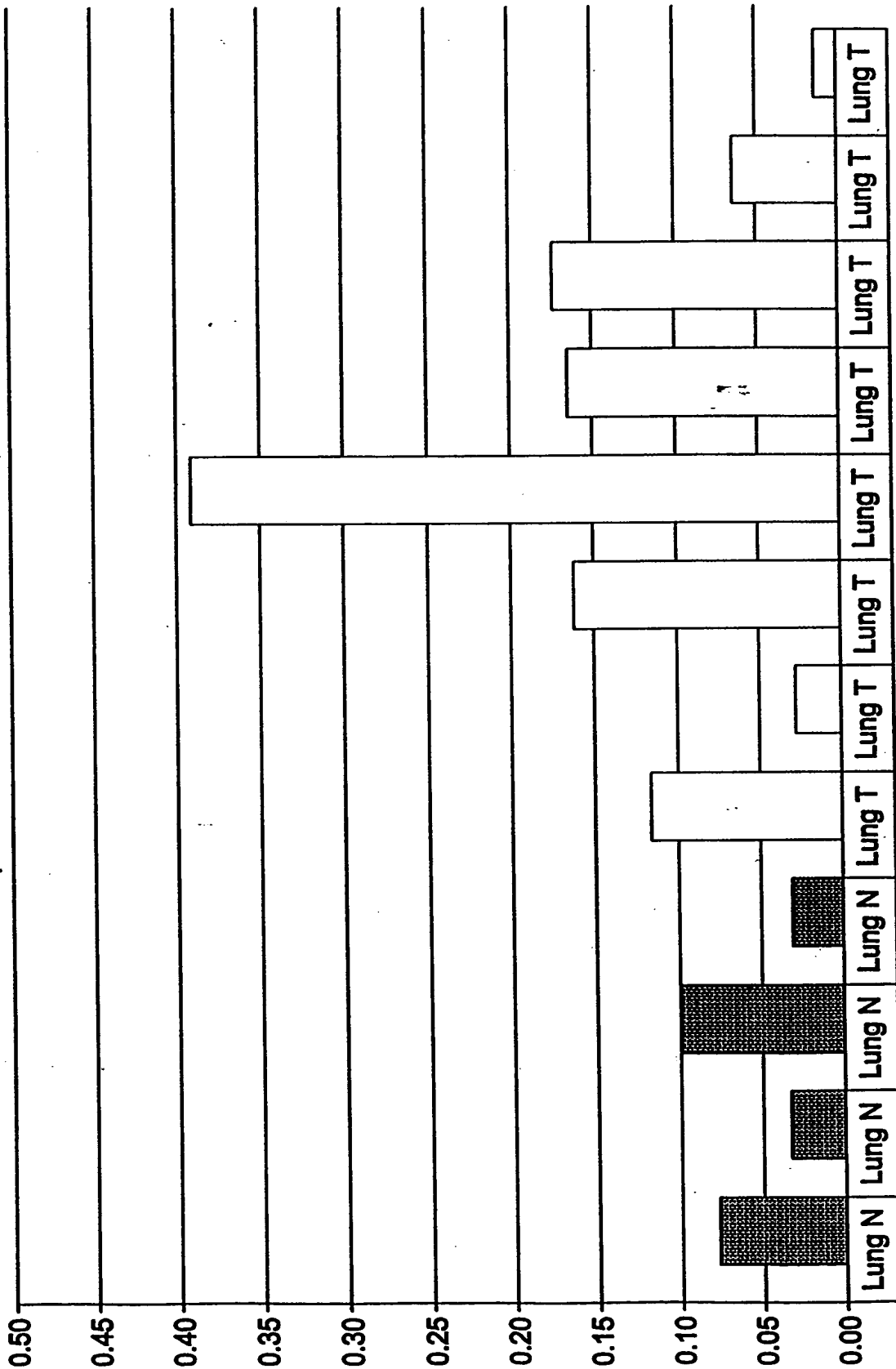
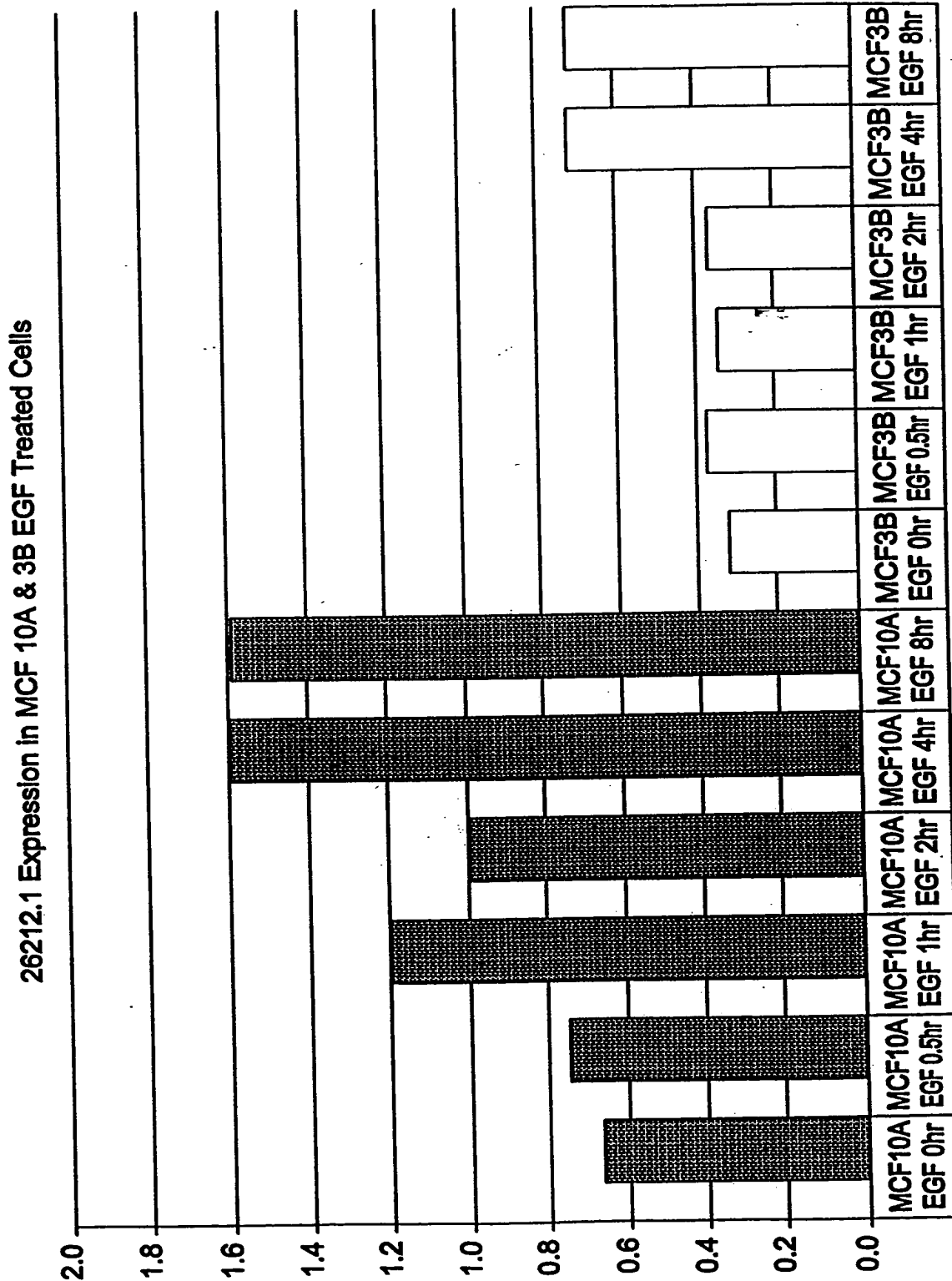
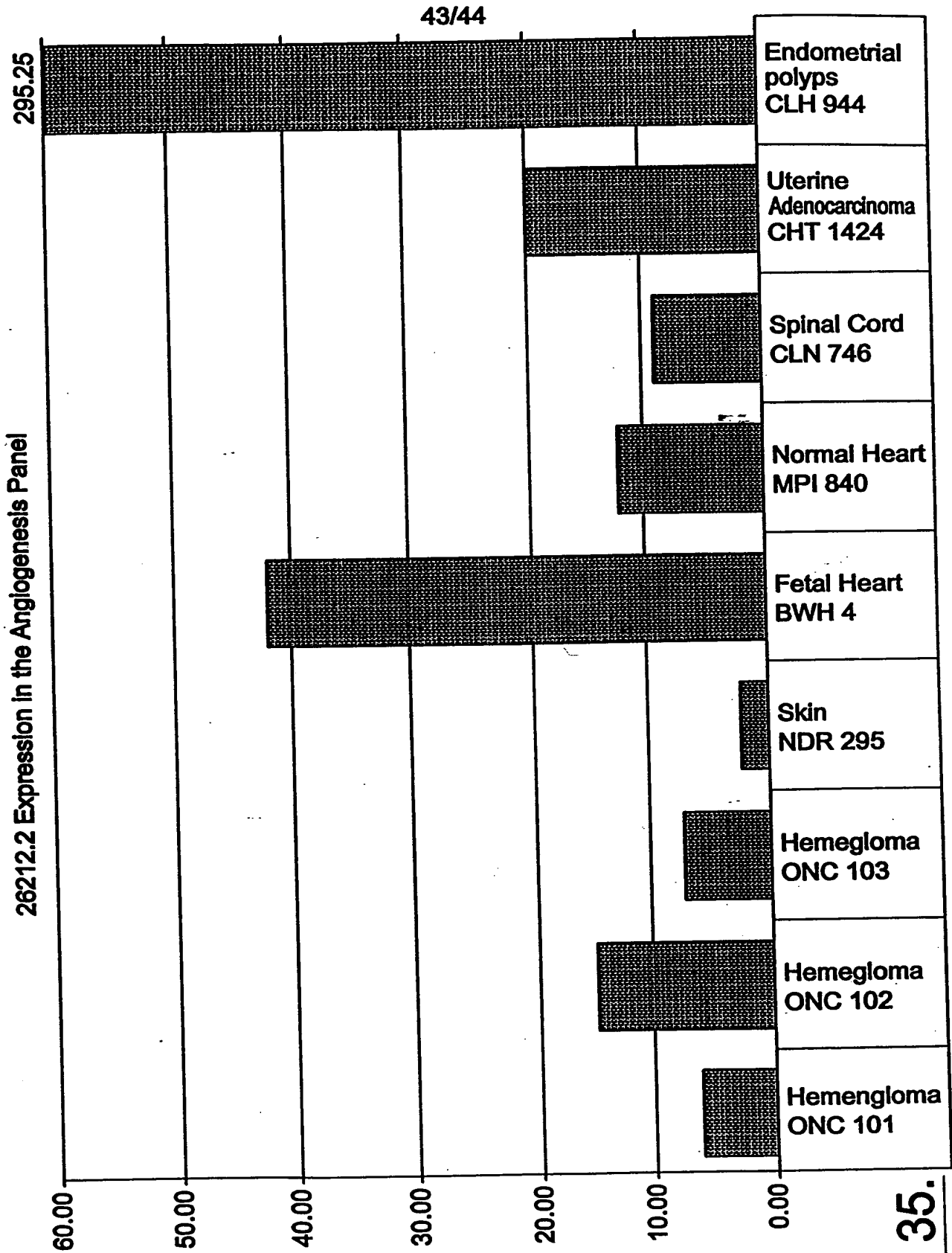


FIG. 33.



**FIG. 34.**



**FIG. 35.**

44/44

FIG. 36.

